GENERALIZED SEMIPARAMETRIC VARYING-COEFFICIENT MODELS FOR LONGITUDINAL DATA

by

Li Qi

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2015

Approved by:

Dr. Yanqing Sun

Dr. Jiancheng Jiang

Dr. Weihua Zhou

Dr. Donna M. Kazemi

ABSTRACT

LI QI. Generalized semiparametric varying-coefficient models for longitudinal data.
(Under the direction of DR. YANQING SUN)

In this dissertation, we investigate the generalized semiparametric varying-coefficient models for longitudinal data that can flexibly model three types of covariate effects: time-constant effects, time-varying effects, and covariate-varying effects, i.e., the covariate effects that depend on other possibly time-dependent exposure variables.

First, we consider the model that assumes the time-varying effects are unspecified functions of time while the covariate-varying effects are parametric functions of an exposure variable specified up to a finite number of unknown parameters. Second, we consider the model in which both time-varying effects and covariate-varying effects are completely unspecified functions. The estimation procedures are developed using multivariate local linear smoothing and generalized weighted least squares estimation techniques. The asymptotic properties of the proposed estimators are established. The simulation studies show that the proposed methods have satisfactory finite sample performance. ACTG 244 clinical trial of HIV infected patients are applied to examine the effects of antiretroviral treatment switching before and after HIV developsing the 215-mutation. Our analysis shows benefit of treatment switching before developing the 215-mutation.

The proposed methods are also applied to the STEP study with MITT cases showing that they have broad applications in medical research.

ACKNOWLEDGMENTS

Upon the completion of this thesis I would sincerely gratefully express my thanks to many people. I would like to give my deepest gratitude to my dissertation advisor, Dr. Yanqing Sun for her guidance, insights and encouragement throughout my dissertation research process. Her attitude to work and to life deeply engraved in my heart and memory. I am deeply grateful for her financial support as well.

I also would like to thank the committee members, Drs. Jiancheng Jiang, Weihua Zhou, and Donna Kazemi for their constructive comments and valuable suggestions. Not forgetting to all the honorable professors at UNCC who supported me on such an unforgettable and unique study experience for five years. I also thank Dr. Peter Gilbert at Fred Hutchinson Cancer Research Center for helpful discussions and reviewing the manuscript.

This research was partially supported by NIAID NIH award R37AI054165, and by NSF grant DMS-1208978. I would like to thank the ACTG for providing the ACTG 244 data, in particular Ronald Bosch and Justin Ritz for preparing the data set and helpful discussions. I also wish to thank the ACTG 244 study participants and study team, including the study chairs Douglas L. Mayers & Thomas C. Merigan. The ACTG 244 project described was supported by Award Numbers U01 A038855, AI038858, AI068634 and AI068636 from NIAID and supported by NIMH and NIDCR.

In addition, I owe my thanks to my friends and family. I would like to thank my parents who have provided endless support and encouragement. I would also like to thank Haohan Chen for supports and preparation of the presentations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Longitudinal data are common in medical and public health research. In AIDS clinical trials, for example, viral loads and CD4 counts are measured repeatedly during the course of studies. These biomarkers have long been known to be prognostic for both secondary HIV transmission and progression to clinical disease in observational studies (Mellors et al., 1997; HIV Surrogate Marker Collaborative Group, 2000; Quinn et al., 2000; Gray et al., 2001), and more recently in randomized trials (Cohen, 2011). An important objective of the AIDS clinical trials is to examine treatment effectiveness on these longitudinal biomarkers. In this dissertation, we consider new methodologies for analyzing the longitudinal data arising from these studies.

1.1 A Motivating Example

In many medical studies, the treatment of the patients may be switched during the study period or the patients may experience more than one phase of treatment. It is important to understand the temporal effects of the new treatment after switching as well as personalized responses to the switching.

A motivating example is a historical case study of antiretroviral treatment regimens, ACTG 244. Zidovudine (ZDV) was the first drug approved for treatment of HIV infection. Initial approval was based on evidence of a short-term survival advantage over placebo when zidovudine was given to patients with advanced HIV disease.

Shortly after that, zidovudine resistance was associated with disease progression measured by a rise in plasma virus and decline in CD4 cell counts in both children and adults receiving zidovudine monotherapy (Japour, 1995; Principi, 2001). Subsequent studies suggested benefits of switching patients to treatments that combined ZDV with didanosine (ddI) or with ddI plus nevirapine (NVP). ACTG 244 enrolled subjects receivingon ZDV monotherapy and monitored their HIV in plasma bi-monthly for the T215Y/F mutation. When a subject's viral population developed the 215 mutation, the subject was randomized to continue ZDV, add ddI or add ddI plus NVP. A diagram of the longitudinal monitoring times and treatment switching is given in Figure 1. An important question is whether and how the treatment switching has any beneficial effects in treating the HIV infected patients.

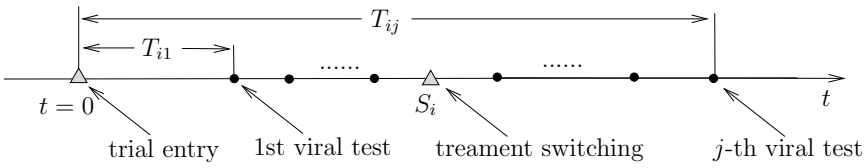


Figure 1: Biomarkers on two time scales: time since the trial entry and time since treatment switching.

The statistical methods developed in this dissertation apply for examining the possible time-varying effects of treatment switching on longitudinal biomarkers such as CD4 cell counts and HIV viral load. These methods further have broad applications since treatment switching is common in medical studies, including switching antiretroviral therapies in response to results of viral load and drug resistance testing (Gilks et al., 2006; Phillips et al., 2008), and, very generally, switching therapies for chronic diseases based on biomarker response results.

We investigate the treatment switching problem under a general semiparametric modelling framework with covariate-varying effects for longitudinal data. In the next section, we will review some of the relevant literature in this area.

1.2 Literature Review

Time-varying semiparametric regression models for longitudinal data have been intensively studied, in which the covariate effects are constant over time for some covariates and time-varying for others. For the semiparametric additive model, the approaches include the nonparametric kernel smoothing by Hoover et al. (1998), the joint modeling of longitudinal responses and sampling times by Martinussen and Scheike (1999, 2000, 2001), Lin and Ying (2001), the backfitting method by Wu and Liang (2004) and the profile kernel smoothing approach by Sun and Wu (2005). Fan et al. (2007) proposed a profile local linear approach by imposing some correlation structure for the longitudinal data for improving efficiency. Fan and Li (2004) considered the profile local linear approach and the joint modelling for partially linear models. Hu and Carroll (2004) showed that for partially linear models, the backfitting is less efficient than the profile kernel method. The proportional means model has been studied by Lin and Carroll (2000), Sun and Wei (2000), Cheng and Wei (2000), Hu et al. (2003) and Sun (2010). The generalized linear model with a known link function was studied by Lin and Carroll (2001) using profile-based generalized estimating equations (GEE) and a local linear approach. Lin et al. (2007) proposed a local linear GEE method when all regression coefficients are nonparametric functions of time. Sun et al. (2013b) proposed a profile kernel estimation procedure for the

generalized semiparametric model with time-varying effects without having to specify a sampling model for the observation times and thus avoiding misspecification of the sampling model.

However, in many applications the covariate effects may not only vary with time, but with an exposure variable. The covariate-varying effects models have been widely studied for cross-sectional data since the seminal paper of Hastie and Tibshirani (1993). For cross-sectional survival data, Scheike (2001) proposed a generalized additive Aalen model with two time-scales where, for example, one time-scale is the duration of illness since diagnosis and the other time-scale is the age when the transition to the illness stage occurred. The model allows examining the nonlinear interactions between the covariates and the age. The proportional hazards model with covariate-varying effects was studied by Fan et al. (2006), with an application to the nursing home data where the duration of nursing home stay and the age of residents are the two time-scales. The nonlinear interactions between covariates with an exposure variable were estimated using the local partial-likelihood technique. This approach was extended to model multivariate failure data by Cai et al. (2007, 2008). Yin et al. (2008) studied a partially linear additive hazards regression with varying-covariate effects in which the nonlinear interactions were estimated using the local score function. Chen et al. (2013) studied time-varying effects for overdispersed recurrent event data with treatment switching using spline method. However, we are not aware of any research on the longitudinal models with covariate-varying effects. The motivating examples testify the importance of such development.

CHAPTER 2: SEMIPARAMETRIC MODEL WITH PARAMETRIC COVARIATE-VARYING EFFECTS

2.1 Model

Suppose that there is a random sample of n subjects and τ is the end of follow-up. Let $X_i(t)$ and $U_i(t)$ be possibly time-dependent covariates for the ith subject. Suppose that observations of response process $Y_i(t)$ for subject i are taken at the sampling time points $0 \leq T_{i1} < T_{i2} < \cdots < T_{in_i} \leq \tau$, where n_i is the total number of observations for subject i. The sampling times can be irregular and dependent on covariates. In addition, some subjects may drop out of the study early. Let $N_i(t) = \sum_{j=1}^{n_i} I(T_{ij} \leq t)$ be the number of observations taken from the ith subject by time t, where $I(\cdot)$ is the indicator function. Let C_i be the end of follow-up time or censoring time whichever comes first. The responses for subject i can only be observed at time points before C_i. Thus $N_i(t)$ can be written as $N_i^*(t \wedge C_i)$, where $N_i^*(t)$ is the counting process of sampling times. Assume that $\{Y_i(\cdot), X_i(\cdot), U_i(\cdot), N_i(\cdot), i = 1, \cdots, n\}$ are independent identically distributed (iid) random processes. The censoring time C_i is noninformative in the sense that $E\{dN_i^*(t) \mid X_i(t),\, U_i(t), C_i \geq t\} = E\{dN_i^*(t) | X_i(t), U_i(t)\}$ and $E\{Y_i(t)|X_i(t), U_i(t), C_i \geq t\} = E\{Y_i(t)|X_i(t),\, U_i(t)\}$. Assume that $dN_i^*(t)$ is independent of $Y_i(t)$ conditional on $X_i(t)$, $U_i(t)$ and $C_i \geq t$. The censoring time C_i is allowed to depend on $X_i(\cdot)$ and $U_i(\cdot)$.

Let $X_i(t) = (X_{1i}^T(t), X_{2i}^T(t), X_{3i}^T(t))^T$ consist of three parts of dimensions p_1, p_2 and

p_3, respectively, over the time interval $[0, \tau]$. Let $U_i(t)$ be the scalar covariate process with support $\mathcal{U}$. To characterize the treatment switching effects of $X_{3i}(t)$ with respect to $U_i(t)$, we propose the generalized semiparametric varying-coefficient model

$$\mu_i(t) = E\{Y_i(t)|X_i(t), U(t)\} = g^{-1}\{\alpha^T(t)X_{1i}(t) + \beta^T X_{2i}(t) + \gamma^T(U_i(t);\theta)X_{3i}(t)\}, \tag{2.1}$$

for $0 \leq t \leq \tau$, where $g(\cdot)$ is a known link function, $\alpha(\cdot)$ is a p_1-dimensional vector of completely unspecified functions, β is a p_2-dimensional vector of unknown parameters and $\gamma(\cdot;\theta)$ is a p_3-dimensional vector of parametric functions specified up to a finite number of unknown parameters θ. Setting the first component of $X_{1i}(t)$ as 1 gives a nonparametric baseline function. $\gamma(u)$ is the effect of $X_{3i}(t)$ at the covariate level $U_i(t) = u$. In addition, different link functions can be selected to provide a rich family of models for longitudinal data. Both categorical and continuous longitudinal responses can be modelled with appropriately chosen link functions. For example, the identity and logarithm link functions can be used for continuous response variables while the logit link function can be used for binary responses.

For the motivating example the ACTG 244, t is the time since initiation of antiretroviral therapy (ART). It is of interest to know how biomarkers such as viral load and CD4 counts respond to the new treatments. It is natural to assume that the effects of the new treatments depend on the time duration $U_i(t) = t - S_i$ since the switching, where S_i is the time of treatment switching. Letting $X_{3i}(t) = I(t > S_i)$ in (2.1), $\gamma(u)$ represents the change in the conditional mean response at time u after treatment switching adjusting for other covariates $X_{1i}(t)$ and $X_{2i}(t)$. On the other

hand, if we let $X_{3i}(t) = X_{3i}^o(t)I(t > S_i)$ where $X_{3i}^o(t)$ are the indicators for the new treatments after switching, then $\gamma(u)$ are the effects of new treatments starting from testament switching. Note that for each patient we have two time scales involved, one is the time since study entry t and the other one is the time since treatment switching $U_i(t)$. If there is no switching for the ith patient, we let $U_i(t) = 0$.

2.2 Estimation

In this section, we develop the estimation procedure for model (2.1) when $\alpha(t)$ is an unspecified function and $\gamma(u) = \gamma(u, \theta)$ are parametric functions. The proposed approach utilizes the local linear estimation technique which has been shown to be design-adaptive and more efficient in correcting boundary bias than the kernel smoothing approach in nonparametric estimation (cf., Fan and Gijbels (1996)). The technique was applied by Cai and Sun (2003) and Sun et al. (2009b) to develop a local partial likelihood method for the time-varying coefficients in the Cox regression model, and recently by Sun et al. (2013b) for longitudinal data.

At each t_0, let

$$\alpha(t) = \alpha(t_0) + \dot{\alpha}(t_0)(t - t_0) + O((t - t_0)^2)$$

be the first order Taylor expansion of $\alpha(\cdot)$ for $t \in \mathcal{N}_{t_0}$, a neighborhood of t_0, where $\dot{\alpha}(t_0)$ is the derivative of $\alpha(t)$ at $t = t_0$. Denote $\alpha^*(t_0) = (\alpha^T(t_0), \dot{\alpha}^T(t_0))^T$, $X_{1i}^*(t, t - t_0) = X_{1i}(t) \otimes (1, t - t_0)^T$, where $\otimes$ is the Kronecker product. Let $\zeta = (\beta^T, \theta^T)^T$. For $t \in \mathcal{N}_{t_0}$, model (2.1) can be approximated by

$$\tilde{\mu}(t, t_0, \alpha^*(t_0), \zeta | X_i, U_i) = \varphi\{\alpha^{*T}(t_0)X_{1i}^*(t, t - t_0) + \beta^T X_{2i}(t) + \gamma^T(U_i(t), \theta)X_{3i}(t)\}, \tag{2.2}$$

where $\varphi(\cdot) = g^{-1}(\cdot)$ is the inverse function of the link function $g(\cdot)$.

At each t_0 and for fixed ζ, we propose the following local linear estimating function for $\alpha^*(t_0)$:

$$U_\alpha(\alpha^*;\zeta,t_0) = \sum_{i=1}^{n} \int_0^\tau W_i(t)\,[Y_i(t) - \tilde{\mu}(t,t_0,\alpha^*(t_0),\zeta|X_i,U_i)] \times X_{1i}^*(t,t-t_0)K_h(t-t_0)\,dN_i(t), \tag{2.3}$$

where $W_i(t) = W(t,X_i(t),U_i(t))$ is a nonnegative weight process, $K(\cdot)$ is a kernel function, $h = h_n > 0$ is a bandwidth parameter and $K_h(\cdot) = K(\cdot/h)/h$. The solution to the equation $U_\alpha(\alpha^*;\zeta,t_0) = 0$ is denoted by $\tilde{\alpha}^*(t_0,\zeta)$.

Let $\tilde{\alpha}(t,\zeta)$ be the first p_1 components of $\tilde{\alpha}^*(t,\zeta)$. The profile weighted least squares estimator $\hat{\zeta}$ is obtained by minimizing the following profile least squares function $\ell_\zeta(\zeta)$ with respect to ζ, where

$$\ell_\zeta(\zeta) = \sum_{i=1}^{n} \int_{t_1}^{t_2} Q_i(t)[Y_i(t) - \varphi\{\tilde{\alpha}^T(t,\zeta)X_{1i}(t) + \beta^T X_{2i}(t) + \gamma^T(U_i(t),\theta)X_{3i}(t)\}]^2\,dN_i(t), \tag{2.4}$$

where $Q_i(t)$ is a nonnegative weight process that can be different from $W_i(t)$ and $[t_1,t_2] \subset (0,\tau)$ in order to avoid possible instability near the boundary.. The profile estimator for $\alpha(t_0)$ is obtained by $\hat{\alpha}(t_0) = \tilde{\alpha}(t_0,\hat{\zeta})$ through substitution.

The Newton-Raphson iterative method can be used to find the estimator $\hat{\zeta}$ that minimizes (2.4). Taking the derivative of $\ell_\zeta(\zeta)$ with respect to ζ leads to the following

estimating function

$$U_\zeta(\zeta) = \sum_{i=1}^{n} \int_{t_1}^{t_2} W_i(t) \left[Y_i(t) - \varphi\{\tilde{\alpha}^T(t,\zeta) X_{1i}(t) + \eta^T(U_i(t),\zeta) X_{2i}^*(t)\} \right] \times \left\{ \frac{\partial \tilde{\alpha}(t,\zeta)}{\partial \zeta} X_{1i}(t) + \frac{\partial \eta(U_i(t),\zeta)}{\partial \zeta} X_{2i}^*(t) \right\} dN_i(t), \tag{2.5}$$

where $\eta(U_i(t),\zeta) = (\beta^T, \gamma^T(U_i(t),\theta))^T$, $\partial\eta(U_i(t),\zeta)/\zeta = \text{diag}\{I_{p_2}, \partial\gamma^T(U_i(t),\theta)/\partial\theta\}$ and $X_{2i}^*(t) = (X_{2i}^T(t), X_{3i}^T(t))^T$. Here $\frac{\partial\tilde{\alpha}(t,\zeta)}{\partial\zeta}$ be the first p_1 components of $\frac{\partial\alpha^*(t,\zeta)}{\partial\zeta}$ which can be expressed in terms of the partial derivatives of $U_\alpha(\alpha^*;\zeta,t)$ at $\alpha^* = \tilde{\alpha}^*(t,\zeta)$. Specifically, since $U_\alpha(\tilde{\alpha}^*(t,\zeta);\zeta,t) \equiv \mathbf{0}_{2p_1}$, it follows that $\tilde{\alpha}^*(t,\zeta)$ satisfies

$$\left\{ \frac{\partial U_\alpha(\alpha^*;\zeta,t)}{\partial \alpha^*} \frac{\partial \tilde{\alpha}^*(t,\zeta)}{\partial \zeta} + \frac{\partial U_\alpha(\alpha^*;\zeta,t)}{\partial \zeta} \right\} \Bigg|_{\alpha^* = \tilde{\alpha}^*(t,\zeta)} = \mathbf{0}_{2p_1}.$$

Therefore,

$$\frac{\partial \tilde{\alpha}^*(t,\zeta)}{\partial \zeta} = -\left\{ \frac{\partial U_\alpha(\alpha^*;\zeta,t)}{\partial \alpha^*} \right\}^{-1} \frac{\partial U_\alpha(\alpha^*;\zeta,t)}{\partial \zeta} \Bigg|_{\alpha^* = \tilde{\alpha}^*(t,\zeta)}, \tag{2.6}$$

where

$$\frac{\partial U_\alpha(\alpha^*;\zeta,t_0)}{\partial \alpha^*} = -\sum_{i=1}^{n} \int_0^\tau W_i(t) \dot{\varphi}\{\alpha^{*T}(t_0) X_{1i}^*(t, t-t_0) + \eta^T(U_i(t),\zeta) X_{2i}^*(t)\} \times X_{1i}^*(t, t-t_0)^{\otimes 2} K_h(t-t_0)\, dN_i(t), \tag{2.7}$$

and

$$\frac{\partial U_\alpha(\alpha^*;\zeta,t_0)}{\partial \zeta} = -\sum_{i=1}^{n} \int_0^\tau W_i(t) \dot{\varphi}\{\alpha^{*T}(t_0) X_{1i}^*(t, t-t_0) + \eta^T(U_i(t),\zeta) X_{2i}^*(t)\} \times X_{1i}^*(t, t-t_0) \{ \frac{\partial \eta(U_i(t);\zeta)}{\partial \zeta} X_{2i}^*(t) \}^T K_h(t-t_0)\, dN_i(t). \tag{2.8}$$

When link function is identity function, $\tilde{\alpha}^*(t_0,\zeta)$ can be solved explicitly as the

root of the estimating function (2.3). Under a general link function, $\tilde{\alpha}^*(t_0, \zeta)$ can be solved using the Newton-Raphson iterative algorithm. The estimation procedure iteratively updates estimates of the nonparametric component $\tilde{\alpha}^*(t_0, \zeta)$ and the parametric component ζ. Specifically, the estimators $\hat{\alpha}(t_0)$ and $\hat{\zeta}$ can be accomplished through the following iterated algorithm:

Computational algorithm

1. Given $\hat{\alpha}(t)^{\{0\}}$ and $\hat{\zeta}^{\{0\}}$ as the initial values;

2. For each jump point of $\{N_i(\cdot), i = 1, \cdots, n\}$, say t_0 , the mth step estimator $\hat{\alpha}^{*\{m\}}(t_0) = \tilde{\alpha}^*(t_0, \hat{\zeta}^{\{m-1\}})$ is the root of the estimating function (2.3) satisfying $U_\alpha(\hat{\alpha}^{*\{m\}}(t_0), \hat{\zeta}^{\{m-1\}}, t_0) = 0$, where $\hat{\zeta}^{\{m-1\}}$ is the estimate of ζ at the $(m-1)$th step.

3. The mth step estimator $\hat{\zeta}^{\{m\}}$ is the minimizer of (2.4) obtained after replacing $\tilde{\alpha}(t, \zeta)$ with $\hat{\alpha}^{\{m\}}(t)$, where $\hat{\alpha}^{\{m\}}(t)$ is the first p_1 components of $\hat{\alpha}^{*\{m\}}(t)$.

4. Repeating step 2 and 3, the estimators $\hat{\alpha}^{*\{m\}}(t_0)$ and $\hat{\zeta}^{\{m\}}$ are updated at each iteration until convergence. $\hat{\zeta}$ and $\hat{\alpha}(t_0)$ are $\hat{\zeta}^{\{m\}}$ and the first p_1 components of $\hat{\alpha}^{*\{m\}}(t_0)$, respectively, at convergence.

2.3 Asymptotic Properties

Let ζ_0 and $\alpha_0(t)$ be the true values of ζ and $\alpha(t)$ under model (2.1). Let

$$\mu_i(t) = \varphi\{\alpha_0^T(t)X_{1i}(t) + \eta^T(U_i(t), \zeta_0)X_{2i}^*(t)\},$$

$$\dot{\mu}_i(t) = \dot{\varphi}\{\alpha_0^T(t)X_{1i}(t) + \eta^T(U_i(t), \zeta_0)X_{2i}^*(t)\},$$

and $\epsilon_i(t) = Y_i(t) - \mu_i(t)$. Let $\omega(t)$ be the deterministic limit of $W(t)$ in probability as $n \to \infty$. Define

$$e_{11}(t) = E\big[\omega_i(t)\dot{\mu}_i(t)X_{1i}(t)^{\otimes 2}\lambda_i(t)\xi_i(t)\big],$$

and

$$e_{12}(t) = E\big[\omega_i(t)\dot{\mu}_i(t)X_{1i}(t)\left(\frac{\partial\eta(U_i(t),\zeta)}{\partial\zeta}X_{2i}^*(t)\right)^T\lambda_i(t)\xi_i(t)\big],$$

where $\xi_i(t) = I(C_i \geq t)$.

Let

$$\hat{\mu}_i(t) = \varphi\{\hat{\alpha}^T(t)X_{1i}(t) + \eta^T(U_i(t),\hat{\zeta})X_{2i}^*(t)\},$$

$$\hat{\dot{\mu}}_i(t) = \dot{\varphi}\{\hat{\alpha}^T(t)X_{1i}(t) + \eta^T(U_i(t),\hat{\zeta})X_{2i}^*(t)\},$$

and $\hat{\epsilon}_i(t) = Y_i(t) - \hat{\mu}_i(t)$. Let

$$\hat{E}_{11}(t) = n^{-1}\sum_{i=1}^{n}\int_0^\tau W_i(s)\hat{\dot{\mu}}_i(s)X_{1i}(s)^{\otimes 2}K_h(s-t)\,dN_i(s),$$

and

$$\hat{E}_{12}(t) = n^{-1}\sum_{i=1}^{n}\int_0^\tau W_i(s)\hat{\dot{\mu}}_i(s)X_{1i}(s)\left(\frac{\partial\eta(U_i(t),\zeta)}{\partial\zeta}X_{2i}^*(s)\right)^T K_h(s-t)\,dN_i(s).$$

The following theorem characterizes the asymptotic properties of the proposed estimator $\hat{\zeta}$.

Theorem 2.1. *Under Condition I in the Appendix, the estimator $\hat{\zeta}$ is consistent for ζ_0, and $\sqrt{n}(\hat{\zeta} - \zeta_0)$ converges in distribution to a mean zero Gaussian random vector with covariance matrix $A^{-1}\Sigma A^{-1}$, where*

$$A = E\left[\int_{t_1}^{t_2}\omega_i(t)\dot{\mu}_i(t)\{-(e_{12}(t))^T(e_{11}(t))^{-1}X_{1i}(t) + \tfrac{\partial\eta(U_i(t),\zeta_0)}{\partial\zeta}X_{2i}^*(t)\}^{\otimes 2}\,dN_i(t)\right],$$

and

$$\Sigma = E\left[\int_{t_1}^{t_2} \omega_i(t)\epsilon_i(t)\{-(e_{12}(t))^T(e_{11}(t))^{-1}X_{1i}(t) + \tfrac{\partial\eta(U_i(t),\zeta_0)}{\partial\zeta}X_{2i}^*(t)\}\, dN_i(t)\right]^{\otimes 2}. \quad (2.9)$$

The matrix A can be consistently estimated by

$$\hat{A} = n^{-1}\sum_{i=1}^n \int_{t_1}^{t_2} W_i(t)\hat{\dot{\mu}}_i(t)\{-(\hat{E}_{12}(t))^T(\hat{E}_{11}(t))^{-1}X_{1i}(t) + \tfrac{\partial\eta(U_i(t),\hat{\zeta})}{\partial\zeta}X_{2i}^*(t)\}^{\otimes 2}\, dN_i(t)$$

and Σ can be consistently estimated by

$$\hat{\Sigma} = n^{-1}\sum_{i=1}^n \left(\int_{t_1}^{t_2} W_i(s)\hat{\epsilon}_i(t)\{(\hat{E}_{12}(t))^T(\hat{E}_{11}(t))^{-1}X_{1i}(t) + \tfrac{\partial\eta(U_i(t),\hat{\zeta})}{\partial\zeta}X_{2i}^*(t)\}\, dN_i(s)\right)^{\otimes 2}.$$

Next, we state the asymptotic result for the proposed local estimator $\hat{\alpha}(t)$. Denote $\dot{\alpha}_0(t)$, $\ddot{\alpha}_0(t)$ the first and second derivatives of true $\alpha_0(t)$ with respect to t, respectively.

Theorem 2.2. *Under Condition I in the Appendix, we have that $\hat{\alpha}(t)$ converges to $\alpha_0(t)$ uniformly in $t \in [t_1, t_2]$, and*

$$\sqrt{nh}(\hat{\alpha}(t) - \alpha_0(t) - \frac{1}{2}\mu_2 h^2\ddot{\alpha}_0^T(t)) \xrightarrow{\mathcal{D}} N\left(0, \Sigma_\alpha(t)\right), \quad (2.10)$$

where $\mu_2 = \int_{-1}^1 t^2K(t)\, dt$, $\Sigma_\alpha(t) = (e_{11}(t))^{-1}\Sigma_e(t)(e_{11}(t))^{-1}$, and

$$\Sigma_e(t) = \lim_{n\to\infty} hE\{\textstyle\int_0^\tau \omega_i(s)\epsilon_i(s)X_{1i}(s)K_h(s-t)\, dN_i(s)\}^{\otimes 2}.$$

The variance-covariance matrix $\Sigma_\alpha(t)$ can be estimated consistently replacing $e_{11}(t)$ by $\hat{E}_{11}(t)$ and $\Sigma_e(t)$ by $\hat{\Sigma}_e(t) = n^{-1}\sum_{i=1}^n \{\hat{g}_i(t)\}^{\otimes 2}$, where

$$\begin{aligned}\hat{g}_i(t) =& h^{1/2}\int_0^\tau W_i(s)K_h(s-t)X_{1i}(s)\hat{\epsilon}_i(s)\, dN_i(s) - h^{1/2}\hat{E}_{12}(t)\\ & \hat{A}^{-1}\int_{t_1}^{t_2} W_i(s)\hat{\epsilon}_i(s)\left\{\frac{\partial\eta(U_i(t),\hat{\zeta})}{\partial\zeta}X_{2i}^*(s) - (\hat{E}_{12}(s))^T(\hat{E}_{11}(s))^{-1}X_{1i}(s)\right\} dN_i(s).\end{aligned}$$

Note that the estimation error of ζ does not appear in the asymptotic variance of $\alpha(t)$, however for small samples, we suggest estimate high order terms to get better performance of coverage probabilities.

Let $\alpha^{(k)}(t)$ be the k-th component of $\alpha(t)$. Similar notations are used throughout with the superscript (k) denoting the kth component of the corresponding vector. Based on Theorem 2.2, an asymptotic $(1-\alpha)$ pointwise confidence intervals for $\alpha^{(k)}(t)$, $0 < t < \tau$, is obtained by

$$\hat{\alpha}^{(k)}(t) \pm (nh)^{-1/2} z_{\alpha/2} \left[n^{-1} \sum_{i=1}^{n} \{\hat{g}_i^{(k)}(t)\}^2 \right]^{1/2}. \tag{2.11}$$

Let $\mathcal{A}_0(t) = \int_{t_1}^{t} \alpha_0(s)\, ds$ and $\hat{\mathcal{A}}(t) = \int_{t_1}^{t} \hat{\alpha}(s)\, ds$. The following theorem presents a weak convergence result for $G_n(t) = n^{1/2}(\hat{\mathcal{A}}(t) - \mathcal{A}_0(t))$ over $t \in [t_1, t_2]$. This result provides theoretical justifications for construction of simultaneous confidence bands of $\mathcal{A}(t) = \int_{t_1}^{t} \alpha(s)\, ds$ developed later.

Theorem 2.3. *Under Condition I,* $G_n(t) = n^{-1/2} \sum_{i=1}^{n} H_i(t) + o_p(1)$ *uniformly in* $t \in [t_1, t_2] \subset (0, \tau)$, *where*

$$\begin{aligned} H_i(t) \;=\; & \int_{t_1}^{t} (e_{11}(s))^{-1} \int_0^{\tau} \omega_i(u) K_h(u-s) X_{1i}(u) \{Y_i(u) - \mu_i(u)\}\, dN_i(u)\, ds \\ & - \int_{t_1}^{t} (e_{11}(s))^{-1} e_{12}(s)\, ds\, A^{-1} \\ & \times \int_{t_1}^{t_2} \omega_i(s) \{ \frac{\partial \eta(U_i(t), \zeta)}{\partial \zeta} X_{2i}^*(s) - (e_{12}(s))^T (e_{11}(s))^{-1} X_{1i}(s) \} \\ & \times \{Y_i(s) - \mu_i(s)\}\, dN_i(s). \end{aligned} \tag{2.12}$$

The processes $G_n(t)$ *converges weakly to a zero-mean Gaussian process on* $[t_1, t_2]$.

The asymptotic covariance matrix of $G_n(t)$ can be estimated consistently by $\hat{\Sigma}_G(t) =$

$n^{-1}\sum_{i=1}^{n}\{\hat{H}_i(t)\}^{\otimes 2}$, where $\hat{H}_i(t)$ is obtained by replacing the unknown quantities in $H_i(t)$ with their corresponding empirical counterparts.

For time-varying coefficient models, simultaneous confidence bands for the estimated coefficient functions are more desirable than the pointwise confidence intervals. Motivated by the Gaussian multiplier resampling method of Lin et al. (1993), we define $G_n^*(t) = n^{-1/2}\sum_{i=1}^{n}\hat{H}_i(t)\phi_i$, where $\phi_1, \phi_2, \ldots, \phi_n$ are iid standard normal random variables independent from the observed data set. By Lemma 1 of Sun and Wu (2005), the processes $G_n(t)$ and $G_n^*(t)$ given the observed data sequence converge weakly to the same zero-mean Gaussian process on $[t_1, t_2]$. The distribution of $\sup_{t_1\le t\le t_2}|G_n^{*(k)}(t)/[\sum_{i=1}^{n}\{\hat{H}_i^{(k)}(t)\}^2/n]^{1/2}|$ can be approximated by repeatedly generating samples of $\phi_1, \phi_2, \ldots, \phi_n$. Let $c_{\alpha k}$ be the $100(1-\alpha)$-th the percentile of this approximate distribution, the $(1-\alpha)$ simultaneous confidence bands for $\{\mathcal{A}^{(k)}(t), t \in [t_1, t_2]\}$ is given by

$$\hat{\mathcal{A}}^{(k)}(t) \pm n^{-1/2}c_{\alpha k}\left[n^{-1}\sum_{i=1}^{n}\{\hat{H}_i^{(k)}(t)\}^2\right]^{1/2}.$$

2.4 Bandwidth Selection

The optimal bandwidth h can be selected by minimizing the asymptotic mean integrated square error (MISE). It follows from Theorem 2.2 that the MISE for the kth component of $\alpha(t)$ is

$$\int_{t_1}^{t_2}\left[\frac{1}{4}h^4\mu_2^2\{\ddot{\alpha}_0^{(k)}(t)\}^2 + \frac{1}{nh}\sigma^{(k)}(t)\right]dt,$$

where $\ddot{\alpha}_0^{(k)}(t)$ is the kth element of the vector $\ddot{\alpha}_0(t)$ and $\sigma^{(k)}(t)$ is the $k$the diagonal element of $(e_{11}(t))^{-1}\Sigma_\alpha(t)(e_{11}(t))^{-1}$. Therefore, the asymptotic optimal bandwidth is

giving by

$$h_{opt,k} = \left[\frac{\int_{t_1}^{t_2} \sigma^{(k)}(t)\, dt}{\int_{t_1}^{t_2} \mu_2^2 \{\ddot{\alpha}_0^{(k)}(t)\}^2\, dt} \right]^{1/5} n^{-1/5}.$$

The optimal theoretical bandwidth is difficult to achieve since it would involve estimating the second derivative $\ddot{\alpha}(t)$; see Fan and Gijbels (1996), Cai and Sun (2003), Sun and Gilbert (2012) and Sun et al. (2013b). In practice, the appropriate bandwidth selection can be based on a cross-validation method. This approach is widely used in the nonparametric function estimation literature; see Rice and Silverman (1991) for a leave-one-subject-out cross-validation approach, and Tian et al. (2005) for a K-fold cross-validation approach for survival data. Sun et al. (2013b) extended the K-fold cross-validation approach for longitudinal data.

An analog of the K-fold cross-validation approach in the current setting is to divide the data into K equal-sized groups. With D_k denoting the kth subgroup of data, the kth prediction error is given by

$$PE_k(h) = \sum_{i \in D_k} \left\{ \int_{t_1}^{t_2} W_i(t)[Y_i(t) - \varphi\{\hat{\alpha}_{(-k)}^T(t) X_{i1}(t) + \eta^T(U_i(t), \hat{\zeta}_{(-k)}) X_{2i}^*(t)\}] dN_i(t) \right\}^2, \tag{2.13}$$

for $k = 1, \cdots, K$, where $\hat{\alpha}_{(-k)}(t)$ and $\hat{\zeta}_{(-k)}$ are estimated using the data from all subgroups other than D_k. The data-driven bandwidth selection based on the K-fold cross-validation is obtained by minimizing the total prediction error $PE(h) = \sum_{k=1}^{K} PE_k(h)$, with respect to h.

We also investigate another user friendly bandwidth selection method. The bandwidth selection formula $h = C\hat{\sigma}_w n^{-1/3}$ has been examined for the nonparametric density estimation and for semiparametric failure time regression in Jones et al. (1991),

Zhou and Wang (2000) and Sun et al. (2013a) among others, where C is a constant and $\hat{\sigma}_w$ is the estimated standard error of the sampling times in the domain of the nonparametric functions to be estimated. To adopt the formula for the longitudinal data, we note that the observation times $\{T_{ij}, j = 1, \ldots, n_i\}$ for a subject i are likely dependent. Suppose that ϕ_i is the random effect that induce such dependence. Then variance of the observation times can be expressed as

$$\sigma_T^2 = Var(T_{ij}) = E\{Var(T_{ij}|\phi_i)\} + \text{Var}\{E(T_{ij}|\phi_i)\}.$$

The σ_T^2 can be estimated by $\hat{\sigma}_T^2 = n^{-1}\sum_{i=1}^{n} S_i^a + S^b$, where S_i^a is the sample variance of T_{ij}, for $j = 1, \ldots, n_i$, for the ith subject, and S^b is the sample variance of $\bar{T}_{i\cdot} = n_i^{-1}\sum_{i=1}^{n_i} T_{ij}$.

Our simulation shows that the K-fold cross-validation method works well for K is the range of 3 to 10. The constant C in the bandwidth selection formula $h = C\hat{\sigma}_T n^{-1/3}$ is calibrated through simulations under different models and sample sizes, where the cross-validation selected bandwidths h_{CV} are taken as the responses. Some of the results are summarized in Table 1. Although C can be different for different settings, it falls in the range between 3 and 5. The estimation and hypothesis testing results are not very sensitive to the choice of $C \in [3, 5]$. In practice, a larger C can be used if the distribution of the sampling times is skewed or sparse in some areas.

2.5 Weight Function Selection

Since the data used in (2.3) are localized in a neighborhood of t, a weight function for (2.3) will not have much effect on the local linear estimator for $\alpha(t)$.(Sun

et al., 2013b). And under Theorem 2.1, the proposed estimator $\hat{\zeta}$ is consistent and asymptotically normal as long as the weight process $W(\cdot)$ converges in probability to a deterministic function $\omega(\cdot)$. However the selection of $W(\cdot)$ does affect the efficiency of the estimator $\hat{\zeta}$. Naturally, we would like to choose the optimal weight such that the asymptotic variance of $\hat{\zeta}$ is minimized. Wang et al. (2005) showed that, in semiparametric setting, the estimator for the parametric component in the model will achieve the semiparametric efficiency bound only if the within-cluster correlation matrix is specified correctly. One typically takes the weight matrix in the weighted least squares method to be the inverse of estimated covariance matrix. Challenges arise in estimating the covariance function nonparametrically due to the fact that longitudinal data are frequently collected at irregular and possibly subject-specific time points.

When the link function is identity, Wu and Pourahmadi (2003) proposed nonparametric estimation of large covariance matrices using a two-step estimation procedure (Fan and Zhang, 2000), but their method can deal with only balanced or nearly balanced longitudinal data. Huang et al. (2006) introduced a penalized likelihood method for estimating a covariance matrix when the design is balanced and Yao et al. (2005a,b) approached the problem from the standpoint of functional data analysis. Fan et al. (2007) proposed a quasi-likelihood approach and they studied the case in which the observation times are irregular on a continuous time interval. In their method, the variance function is modeled nonparametrically as a function in time, but the correlation is assumed to be a member of a known family of parametric correlation functions. Li (2011) extended these methods by modelling the covariance

function completely nonparametrically.

This topic for models with non identity link is beyond the scope of this thesis. Here we suppose that the repeated measurements of $Y_i(\cdot)$ within the same subject are independent and that $Y_i(\cdot)$ is independent of $N_i(\cdot)$ conditional on the covariates $X_i(t)$ and $U_i(t)$. Let $\sigma_\epsilon^2(t|X,U) = \mathrm{Var}\{Y_i(t)|X_i(t),U_i(t)\}$ be the conditional variance of $Y_i(t)$ given $X_i(t)$ and $U_i(t)$. Then

$$\Sigma = E\left[\int_{t_1}^{t_2} \omega_i^2(t)\sigma_\epsilon^2(t)\{-(e_{11}(t))^{-1}e_{12}(t)X_{1i}(t) + \frac{\partial\eta(U_i(t),\zeta_0)}{\partial\zeta}X_{2i}^*(t)\}^{\otimes 2}\xi_i(t)\lambda_i(t)\right].$$

When $\omega_i(t) = \dot{\mu}_i(t)/\sigma_\epsilon^2(t|X,U)$, it often leads to asymptotically efficient estimators in many semiparametric models discussed by Bickel et al. (1993) and Sun et al. (2013b).

Practically a two-stage estimation procedure can be considered to improve the efficiency of the estimation of ζ. In the first stage, unit weight function is used to obtain $\hat{\alpha}_I(t)$ and $\hat{\zeta}_I$. In the second stage, the updated estimators $\hat{\zeta}_W$ are obtained by choosing the weight $W_i(t) = \hat{\dot{\mu}}_i(t)/\hat{\sigma}_\epsilon^2(t|X,U)$, where $\hat{\dot{\mu}}(t) = \dot{\varphi}\{\hat{\alpha_I}^T(t)X_{1i}(t) + \eta^T(U_i(t),\hat{\zeta}_I)X_{2i}^*(t)\}$. And the nonparametric estimation of the variance process is

$$\hat{\sigma}_\epsilon^2(t) = \frac{\sum_{i=1}^n \int_0^\tau \hat{\epsilon}_i(s)^2 K_h(t-s)dN_i(s)}{\sum_{i=1}^n \int_0^\tau K_h(t-s)dN_i(s)}, \tag{2.14}$$

where $\hat{\epsilon}_i(s) = Y_i(s) - \varphi\{\hat{\alpha_I}^T(t)X_{1i}(t) + \eta^T(U_i(t),\hat{\zeta}_I)X_{2i}^*(t)\}$ is the residuals in the first stage. Note that in logit link for Bernoulli data, $\hat{\sigma}_\epsilon^2(t) = \hat{\mu}_i(t)(1-\hat{\mu}_i(t))$, so the weight function cancels out to be identity.

2.6 Link Function Selection

Our estimation procedure for model (2.1) holds for a wide class of link functions. This presents an opportunity to select the most appropriate link function for a partic-

ular application. In some applications the choice may be based on prior knowledge, but data-driven link function selection is also appealing in many applications. In practice, the link function can be selected to minimize the regression deviation:

$$RD_k(g(\cdot), h_{cv}) = \sum_{i=1}^{n} \left\{ \int_{t_1}^{t_2} W_i(t)[Y_i(t) - \varphi\{\hat{\alpha}_g^T(t)X_{i1}(t) + \eta^T(U_i(t), \hat{\zeta}_g)X_{2i}^*(t)\}]\, dN_i(t) \right\}^2, \tag{2.15}$$

where h_{cv} is the bandwidth selected based on the K-fold cross-validation method for the given link function $g(\cdot)$ described in section 2.4, and $\hat{\alpha}_g(t)$ and $\hat{\zeta}_g$ are the estimators with such bandwidth.

2.7 Simulations

We conducted a simulation study to assess the finite sample performance of the proposed methods. Performance is illustrated under models with three popular link functions below:

$$\text{Identity} : Y_i(t) = \alpha_0(t) + \alpha_1(t)X_{1i} + \beta X_{2i} + (\theta_1 + \theta_2(t - S_i))X_{3i}^*(t) + \epsilon(t);$$

$$\text{Log} : Y_i(t) = \exp\{\alpha_0(t) + \alpha_1(t)X_{1i} + \beta X_{2i} + (\theta_1 + \theta_2(t - S_i))X_{3i}^*(t)\} + \epsilon(t);$$

$$\text{Logit} : \text{logit}\{P(Y_i(t) = 1)\} = \alpha_0(t) + \alpha_1(t)X_{1i} + \beta X_{2i} + (\theta_1 + \theta_2(t - S_i))X_{3i}^*(t),$$

for $0 \le t \le \tau$ with $\tau = 3.5$, where $\alpha_0(t) = 0.5\sqrt{t}$, $\alpha_1(t) = 0.5\sin(t)$ and $\zeta = (\beta, \theta_1, \theta_2) = (0.9, 0.3, -0.6)$, X_{1i} and X_{3i} are uniform random variables on $[-1, 1]$, X_{2i} is a Bernoulli random variable with success probability of 0.5, S_i is a uniform random variable on $[0, 1]$ and $X_{3i}^*(t) = X_{3i}I(t > S_i)$. The error $\epsilon_i(t)$ has a normal distribution with mean ϕ_i and variance 0.5^2, and ϕ_i is $N(0, 1)$. The observation time follows a Poisson process with the proportional mean rate model $h(t|X_i, S_i) = 1.5\exp(0.7X_{2i})$.

The censoring times C_i are generated from a uniform distribution on $[1.5, 8]$. There are approximately six observations per subject on $[0, \tau]$ and about 30% subjects are censored before $\tau = 3.5$. The Epanechnikov kernel $K(u) = 0.75(1 - u^2)I(|u| \leq 1)$ and the unit weight function are used. We take $t_1 = h/2$ and $t_2 = \tau - h/2$ in the estimating functions (2.5) to avoid large variations on the boundaries.

The performances of the estimators for ζ and $\alpha(t)$ at a fixed time t are measured through the Bias, the sample standard error of the estimators (SEE), the sample mean of the estimated standard errors (ESE) and the 95% empirical coverage probability (CP). We take $n = 200, 400$ and 600 and consider bandwidths $h = 0.2$, 0.3, 0.4 and the data-driven selected bandwidth $h_C = 4\hat{\sigma}_T n^{-1/3}$, $h_C = 0.68$ for $n = 200$, $h_C = 0.54$ for $n = 400$ and $h_C = 0.47$ for $n = 600$. Table 2-4 summarize the Bias, SEE, ESE and CP for ζ under the three different link functions, the logarithm link function, the identity link function and the logit link function. Each entry of the tables is calculated based on 500 repetitions. Table 2-4 show that the estimates are unbiased and there is a good agreement between the estimated and empirical standard errors. The bias and variances of the estimates decrease as the sample size increases. The coverage probabilities are close to the 95% nominal level. The simulation results are not overly sensitive to these selection choices. The simulation results also show that the data-driven bandwidth formula $h_C = 4\hat{\sigma}_T n^{-1/3}$ works well.

Figure 2-4 show the plots of the bias, SSE, ESE and the coverage probability for the estimators of $\alpha_0(t)$ and $\alpha_1(t)$ over the time interval $[0, 3.5]$. The plots show that the estimates are close to the true values and the ESE provides a good approximation for the SSE of the pointwise estimators. The empirical coverage probabilities are close

to the 95% nominal level.

Note that our weight function was derived under the independency of the error process in Section 2.5. To evaluate the performance of the weigh function selection method, we did some sensitivity analysis for some different error forms. In Error Model I, the random error $\epsilon_i(t)$ is normally distributed conditional on the ith subject with mean ϕ_i and variance 0.5^2 and ϕ_i is $N(0,1)$. In this model, there are correlations between $\epsilon_i(t)$ and $\epsilon_i(s)$. In Error Model II, $\epsilon_i(t)$ is taken to be a Gaussian process with mean 0, variance function $\sigma_\epsilon^2(t) = 2\sin^2(2t)$. The error process is nonstationary as the variance function is time-dependent. In Error Model III, $\epsilon_i(t)$ has a normal distribution conditional on the ith subject with mean ϕ_i and variance $2\sin^2(2t)$ and ϕ_i is $N(0,1)$. Define the empirical relative efficiency (eff) of the weighted estimator $\hat{\zeta}_W$ to $\hat{\zeta}_I$ as

$$\text{eff}(\zeta) = \left(\frac{\text{SSE}(\hat{\zeta}_\text{I})}{\text{SSE}(\hat{\zeta}_\text{W})} \right)^2 .$$

Larger eff means more efficiency is improved by the introduced weight function.

It's of interest to test whether there are varying effects of X_3 with time since treatment switching, which is to test $H_0 : \theta_2 = 0$. Based on the asymptotic normality of $\hat{\zeta} = (\hat{\beta}^T, \hat{\theta}^T)^T)$ and the estimator for its asymptotic covariance given inTheorem 2.1, the test statistic is taken as $Z = \hat{\theta}_2 / se(\hat{\theta}_2)$, where $se(\hat{\theta}_2)$ is the estimated standard error of $\hat{\theta}_2$. The observed sizes of the test statistic are calculated under the null hypothesis $H_0 : \theta_2 = 0$. The powers of the test are calculated from $\theta_2 = .01$ to .5 by .01. A larger value of θ_2 indicates an increased departure from the null hypothesis. The power curve of the test against θ_2 at 5% nominal level is potted in Figure 5 for

$n = 400$, and bandwidth $h = .2, .3$ and $.4$.

2.8 Application to the ACTG 244 trial

Zidovudine resistance (ZDV^R) has been associated with clinical progression in HIV infected patients. ACTG 244 was a randomized, double-blind trial that evaluated the clinical utility of monitoring for the ZDV^R mutation T215Y/F in HIV reverse transcriptase in asymptomatic HIV infected subjects taking ZDV monotherapy. Subjects' plasma was tested for T215Y/F bi-monthly, and upon detection were randomized to continue ZDV, add didanosine (DDI) or add DDI plus nevirapine (NVP).

The primary objectives of ACTG 244 included: (1) to determine whether a decline in CD4 cell counts was preceded by the T215Y/F mutation and (2) to determine whether initiating alternative antiretroviral regimens based on T215Y/F detection could alter the course of CD4 cell decline associated with clinical failure on ZDV monotherapy.

Among the 289 subjects enrolled, 284 were dispensed ZDV, among whom 57 developed T215Y/F. Forty-nine of these subjects were randomized to ZDV (n=17), ZDV+ddI (n=15), or ZDV+ddI+NVP (n=17), and the other eight subjects went off treatment prior to randomization. Of the 234 treated subjects who were not randomized, 137 remained on ZDV treatment without the T215Y/F mutation until the study was modified after the interim review: 69 (68) subjects were randomized to ZDV+ddI (ZDV+ddI+NVP), and 97 (33.6%) subjects went off treatment prior to the interim review. Table 6 illustrates baseline demographics.

The impact of the 215-mutation based randomization was measured primarily by

the square root CD4 cell count (measured by flow cytometry) and $\log_{10}$plasma HIV RNA. We focus on the CD4 cell count endpoint, which is an independent predictor of AIDS/death (cf. Grabar et al. (2000); Kaufmann et al. (1998); Piketty et al. (1998); Mellors et al. (1997)) and is a partially valid surrogate endpoint (GROUP, 2000). CD4 cell count and T215Y/F mutation status (determined by RT-PCR (Larder et al., 1991)) were measured at study entry and every 8 weeks thereafter, with variability in visit dates across individuals. Figure 6 shows histograms of the visit times and of the first and second randomization times, all since entry into ACTG 244.

2.8.1 Analysis of the Effects of Switching Treatments After Drug-resistant Virus Was Detected

First, we examine the effects of switching treatments following detection of the T215Y/F mutation. Let Y be the square root of CD4 count, Z_1 be Sex (1 if Female; 0 if Male), Z_2 be Age in years at study entry, Z_3 and Z_4 be dummy variables coding race ($Z_3 = 1$ if white and 0 otherwise, $Z_4 = 1$ if black and 0 otherwise), S be the time of the codon 215 mutation, $Trt_{1i}(t) = 1$ if randomized to ZDV and 0 otherwise, $Trt_{2i}(t) = 1$ if randomized to ZDV+ddI and 0 otherwise, and $Trt_{3i}(t) = 1$ if randomized to ZDV+ddI+NVP and 0 otherwise; note that all three indicators are zero prior to detection of the mutation. After preliminary exploration of the data, we propose the following model for each subject i:

$$\begin{aligned} Y_i(t) =& \alpha_0(t) + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 Z_{4i} + (\theta_1 + \theta_2(t - S_i))Trt_{1i}(t) \\ &+ (\theta_3 + \theta_4(t - S_i))Trt_{2i}(t) + (\theta_5 + \theta_6(t - S_i))Trt_{3i}(t) + \epsilon_i(t) \end{aligned} \tag{2.16}$$

for $t \in [0, 2]$. We can use the 3-fold cross-validation method to select the optimal bandwidth. As shown in Figure 7, $h = 0.47$ yielded the smallest prediction error.

We calculate $\hat{\sigma}_T = 0.5923$. By the proposed bandwidth formula, $h = C\hat{\sigma}_T n^{-1/3}$ with $C = 4$, the selected bandwidth is around $h = 0.36$. The results are insensitive to the bandwidths between 0.36 and 0.47.

The time-invariant parameter estimates are presented in Table 7. The estimated baseline function $\hat{\alpha}_0(t)$ with 95% pointwise confidence intervals, the point and 95% pointwise confidence intervals of the switching-treatment effect parameters $\gamma_k(u) = \theta_k + \theta_{k+1}u$ for $k = 1, 3$ and 5 are presented in Figure 8, where u is the time since T215Y/F mutation-based treatment switching according to the first randomization.

The results show that CD4 counts decrease over time, are significantly higher for older individuals, and are not significantly affected by sex and race. None of the treatment effect parameters shows a significant effect. While the estimated $\gamma_k(u)$'s are all not statistically different from zero, they are all negative, indicating that none of the randomly assigned treatments increased CD4 counts after the codon 215 mutation. This analysis suggests lack of benefit of switching treatments (among those available in the study) after drug-resistant virus was detected.

2.8.2 Analysis of the Effects of Switching Treatments Before Drug-resistant Virus Was Detected

Next, we examine the effects of switching treatments *before* drug-resistant virus, the codon 215 mutation, was detected. After independent review of the study data by the Data Safety Monitoring Board in September 1996, all subjects were offered ran-

domization to the ZDV+ddI or ZDV+ddI+NVP arms with six months of additional follow-up.

We let S be the time of the second randomization after interim review and only include subjects without 215 mutation in the analysis. The model is similar to before:

$$\begin{aligned} Y_i(t) =& \alpha_0(t) + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 Z_{4i} \\ &+ (\theta_1 + \theta_2(t - S_i))Trt_{2i}(t) + (\theta_3 + \theta_4(t - S_i))Trt_{3i}(t) + \epsilon_i(t). \end{aligned} \tag{2.17}$$

The results for parameter estimation are in Table 9. From it, $\hat{\theta}_2 = 3.1732$ (p-value=0.0021), $\hat{\theta}_3 = 1.0623$ (p-value=0.0126) and $\hat{\theta}_4 = 2.7819$ (p-value=0.0097). The estimated baseline function $\hat{\alpha}_0(t)$ with 95% pointwise confidence intervals is presented in Figure 9. The estimated switching-treatment effects and 95% confidence bands are above zero, suggesting that ZDV+ddI and ZDV+ddI+NVP improve CD4 counts for patients who have not yet developed the codon 215 drug resistance mutation.

In conclusion, the analyses suggest that switching from ZDV monotherapy to combination therapy improves the CD4 cell count marker of HIV progression for subjects who have not yet had the T215Y/F drug resistance mutation, but treatment switching has little effect after the mutation developed.

Table 1: Average of the cross-validation selected bandwidths, h_{CV}, in 10 repetitions based on 10-fold cross-validation for five different sample sizes and three link functions. The last row of the table includes the values of C calibrated using the formula $h_C = C\sigma_T n^{-1/3}$ under the three models.

	Identity-link	Log-link	Logit-link
n		h_C	
200	0.56	0.52	0.65
300	0.50	0.48	0.90
400	0.41	0.40	0.70
500	0.46	0.35	0.45
600	0.40	0.46	0.33
C	3.31	3.14	4.37

Table 2: Identity-link: Summary of Bias, SEE, ESE and CP for β, θ_1 and θ_2 for different sample sizes and bandwidths. $h_C = 0.68$ for $n = 200$, $h_C = 0.54$ for $n = 400$ and $h_C = 0.47$ for $n = 600$.

		$\beta = .9$				$\theta_1 = .3$				$\theta_2 = -.6$			
n	h	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
200	.2	-.0031	.2705	.2673	.948	-.0050	.1667	.1621	.940	.0004	.0766	.0736	.924
	.3	-.0145	.2775	.2675	.942	.0033	.1650	.1627	.944	-.0062	.0770	.0747	.948
	.4	-.0017	.2857	.2716	.938	.0090	.1695	.1650	.936	.0013	.0800	.0765	.946
	h_C	.0034	.2832	.2729	.930	-.0016	.1782	.1713	.946	-.0020	.0929	.0861	.936
400	.2	.0128	.1987	.1918	.954	.0025	.1234	.1159	.938	-.0009	.0541	.0522	.942
	.3	-.0058	.2045	.1931	.920	.0038	.1197	.1172	.938	-.0045	.0573	.0537	.922
	.4	.0165	.2007	.1927	.944	-.0072	.1163	.1177	.950	.0025	.0564	.0553	.942
	h_C	-.0019	.1972	.1943	.926	.0007	.1219	.1194	.940	.0013	.0608	.0583	.938
600	.2	.0146	.1570	.1572	.938	.0098	.0965	.0954	.948	-.0048	.0420	.0430	.966
	.3	-.0032	.1626	.1577	.942	-.0010	.0986	.0962	.952	.0010	.0456	.0446	.944
	.4	-.0081	.1551	.1584	.958	-.0009	.1032	.0966	.942	-.0011	.0479	.0456	.944
	h_C	-.0059	.1593	.1589	.950	.0025	.1037	.0976	.938	.0005	.0473	.0470	.950

Table 3: Logarithm-link: Summary of Bias, SEE, ESE and CP for β, θ_1 and θ_2 for different sample sizes and bandwidths. $h_C = 0.68$ for $n = 200$, $h_C = 0.54$ for $n = 400$ and $h_C = 0.47$ for $n = 600$.

		$\beta = .9$				$\theta_1 = .3$				$\theta_2 = -.6$			
n	h	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
200	.2	-.0037	.0852	.0813	.936	-.0028	.0569	.0571	.944	.0001	.0255	.0259	.948
	.3	-.0037	.0844	.0824	.952	.0011	.0599	.0573	.948	.0001	.0271	.0264	.936
	.4	.0031	.0818	.0833	.958	-.0027	.0579	.0586	.954	.0006	.0253	.0274	.968
	h_C	-.0040	.0857	.0836	.948	.0007	.0597	.0593	.946	.0003	.0288	.0281	.934
400	.2	-.0016	.0552	.0579	.970	-.0003	.0420	.0407	.950	-.0003	.0188	.0182	.942
	.3	-.0033	.0571	.0584	.956	.0006	.0411	.0411	.938	-.0002	.0189	.0186	.946
	.4	.0006	.0573	.0586	.950	-.0011	.0420	.0413	.940	.0011	.0200	.0190	.936
	h_C	-.0017	.0591	.0589	.956	.0020	.0431	.0418	.938	-.0013	.0202	.0199	.944
600	.2	.0001	.0470	.0475	.950	-.0011	.0331	.0334	.962	.0008	.0152	.0148	.954
	.3	-.0001	.0503	.0479	.940	-.0019	.0348	.0336	.940	.0003	.0155	.0152	.940
	.4	.0009	.0475	.0480	.952	.0007	.0348	.0337	.942	-.0002	.0154	.0156	.960
	h_C	-.0016	.0497	.0482	.940	.0024	.0326	.0341	.960	-.0004	.0158	.0160	.962

Table 4: Logit link: Summary of Bias, SEE, ESE and CP for β, θ_1 and θ_2 for different sample sizes and bandwidths. $h_C = 0.68$ for $n = 200$, $h_C = 0.54$ for $n = 400$ and $h_C = 0.47$ for $n = 600$.

		$\beta = .9$				$\theta_1 = .3$				$\theta_2 = -.6$			
n	h	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP	Bias	SEE	ESE	CP
200	.2	.0147	.2268	.2185	.934	.0233	.2246	.2190	.940	-.0221	.1502	.1450	.948
	.3	.0229	.2108	.2208	.950	.0188	.2377	.2211	.934	-.0206	.1584	.1485	.942
	.4	.0219	.2219	.2204	.948	.0003	.2095	.2223	.960	-.0033	.1449	.1505	.954
	h_C	-.0044	.2452	.2321	.918	.0075	.2388	.2323	.942	-.0126	.1711	.1624	.928
400	.2	-.0083	.1439	.1525	.962	-.0101	.1563	.1531	.942	-.0055	.1067	.1016	.936
	.3	.0110	.1471	.1537	.958	-.0020	.1573	.1544	.944	-.0018	.1063	.1034	.938
	.4	.0083	.1615	.1568	.938	-.0007	.1591	.1571	.952	-.0018	.1079	.1062	.944
	h_C	.0021	.1530	.1602	.960	-.0059	.1614	.1599	.950	.0013	.1130	.1099	.942
600	.2	.0071	.1413	.1363	.934	.0034	.1397	.1374	.940	-.0075	.0916	.0907	.938
	.3	-.0034	.1404	.1376	.954	.0004	.1480	.1386	.930	-.0047	.0996	.0927	.938
	.4	.0025	.1420	.1399	.942	.0067	.1404	.1402	.956	-.0078	.0976	.0948	.950
	h_C	.0014	.1244	.1281	.954	.0049	.1321	.1291	.950	-.0049	.0857	.0879	.952

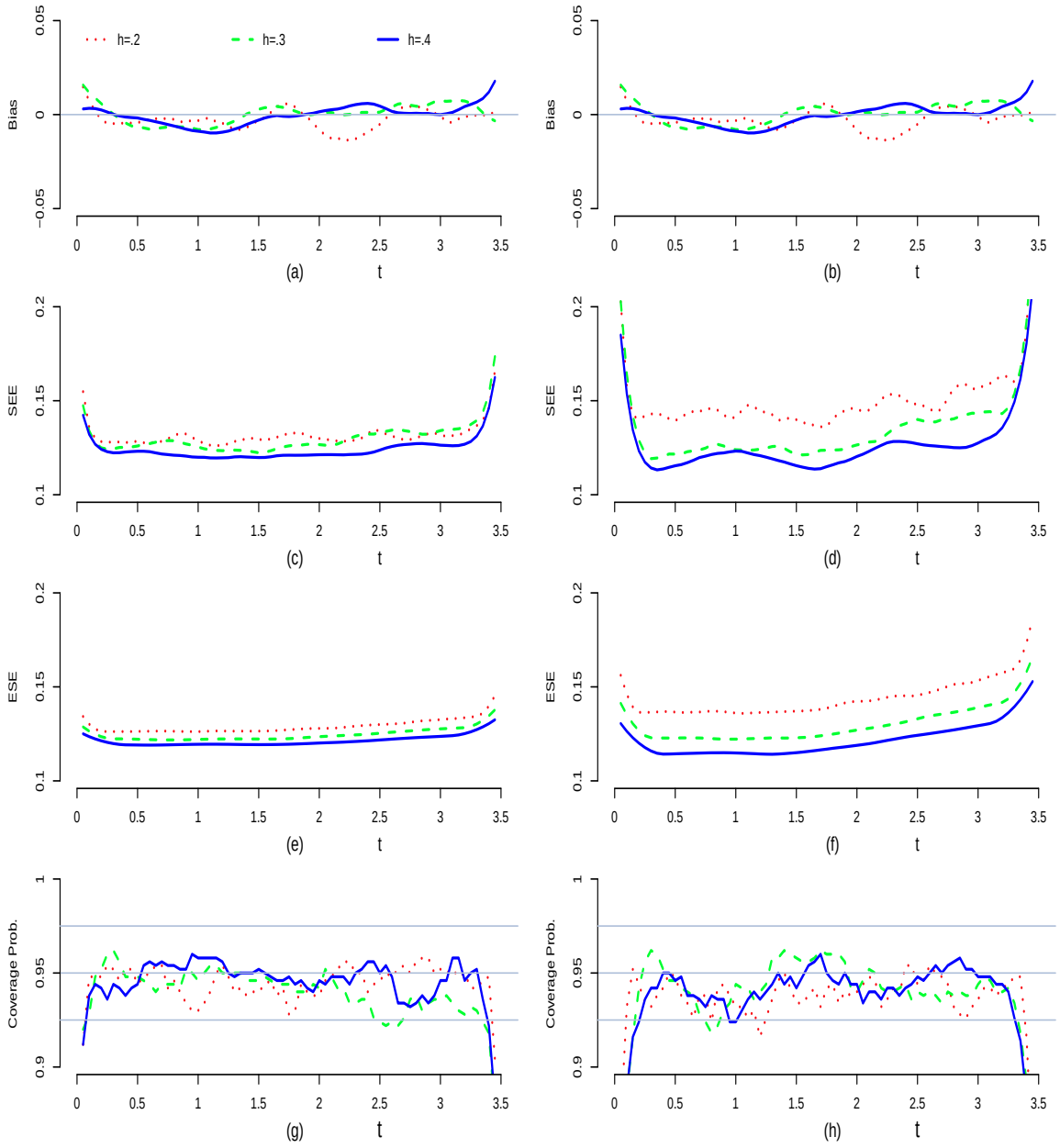


Figure 2: Plots of bias, SEE, ESE and CP for $\hat{\alpha}_0(t)$ and $\hat{\alpha}_1(t)$ under the identity link with n=400. The figures in the left panel are for $\alpha_0(t) = .5\sqrt{t}$, and the figures in the right panel are for $\alpha_1(t) = .5\sin(t)$. Figures (a) and (b) show the bias of $\hat{\alpha}_0(t)$ and $\hat{\alpha}_1(t)$; (c) and (d) show the SSEs; (e), (f) show the ESEs; and (g) and (h) show the CPs based on 500 simulations.

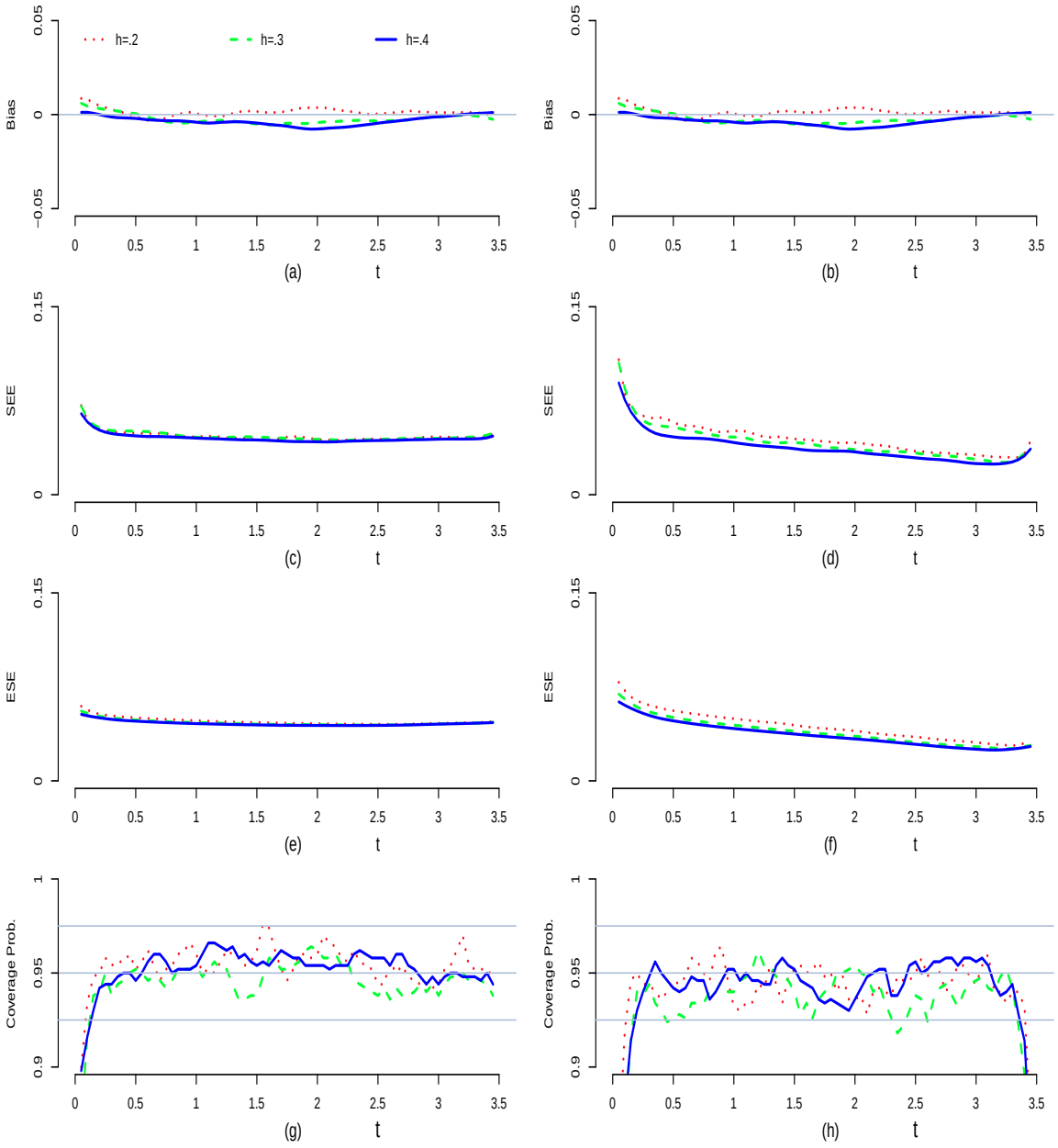


Figure 3: Plots of bias, SEE, ESE and CP for $\hat{\alpha}_0(t)$ and $\hat{\alpha}_1(t)$ under the logarithm link with n=400. The figures in the left panel are for $\alpha_0(t) = .5\sqrt{t}$, and the figures in the right panel are for $\alpha_1(t) = .5\sin(t)$. Figures (a) and (b) show the bias of $\hat{\alpha}_0(t)$ and $\hat{\alpha}_1(t)$; (c) and (d) show the SSEs; (e), (f) show the ESEs; and (g) and (h) show the CPs based on 500 simulations.

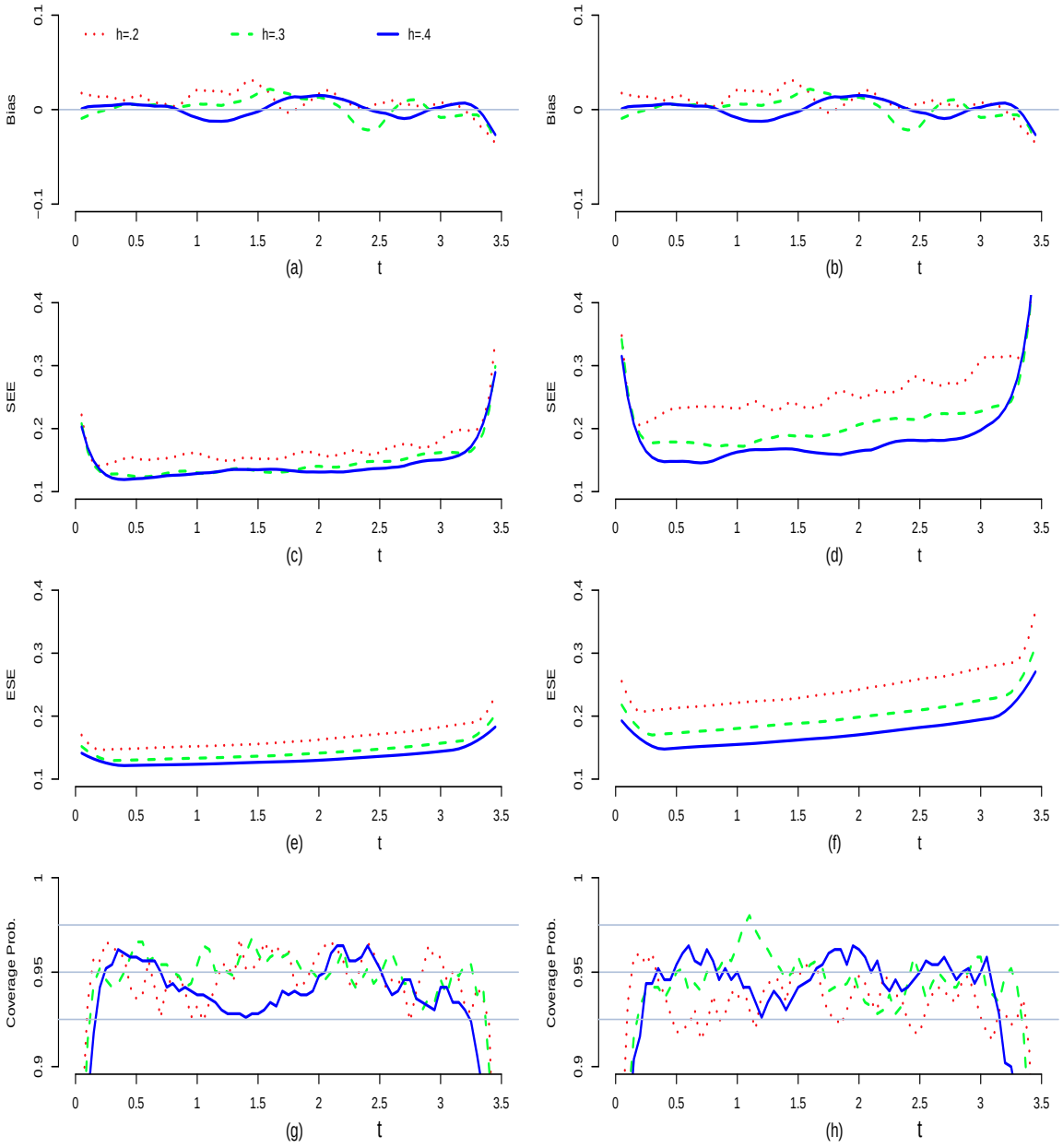


Figure 4: Plots of bias, SEE, ESE and CP for $\hat{\alpha}_0(t)$ and $\hat{\alpha}_1(t)$ under the logit link with n=400. The figures in the left panel are for $\alpha_0(t) = .5\sqrt{t}$, and the figures in the right panel are for $\alpha_1(t) = .5\sin(t)$. Figures (a) and (b) show the bias of $\hat{\alpha}_0(t)$ and $\hat{\alpha}_1(t)$; (c) and (d) show the SSEs; (e), (f) show the ESEs; and (g) and (h) show the CPs based on 500 simulations.

Table 5: The empirical relative efficiency of the estimators of ζ with introduced weight function to the estimators using the unit weight function for n=200.

	Logarithm-link			Identity-link		
h	eff(β)	eff(θ_1)	eff(θ_2)	eff(β)	eff(θ_1)	eff(θ_2)
			Error Model I			
.2	1.3631	1.0373	1.0453	1.0002	1.0093	1.0043
.3	1.4402	1.0246	1.0050	1.0035	1.0017	0.9922
.4	1.2846	0.9676	1.0549	1.0004	1.0039	1.0059
			Error Model II			
.2	3.1695	2.9670	2.1842	2.5538	2.4383	2.6306
.3	2.4600	2.5332	2.2989	1.8118	1.7333	1.9764
.4	2.1410	1.8586	2.0769	1.4311	1.3781	1.6106
			Error Model III			
.2	3.1473	4.4735	5.1587	2.4529	3.4207	4.9425
.3	2.3560	3.1196	3.9545	1.8133	2.1275	2.8485
.4	1.7775	2.3707	2.7515	1.4417	1.4513	1.6595

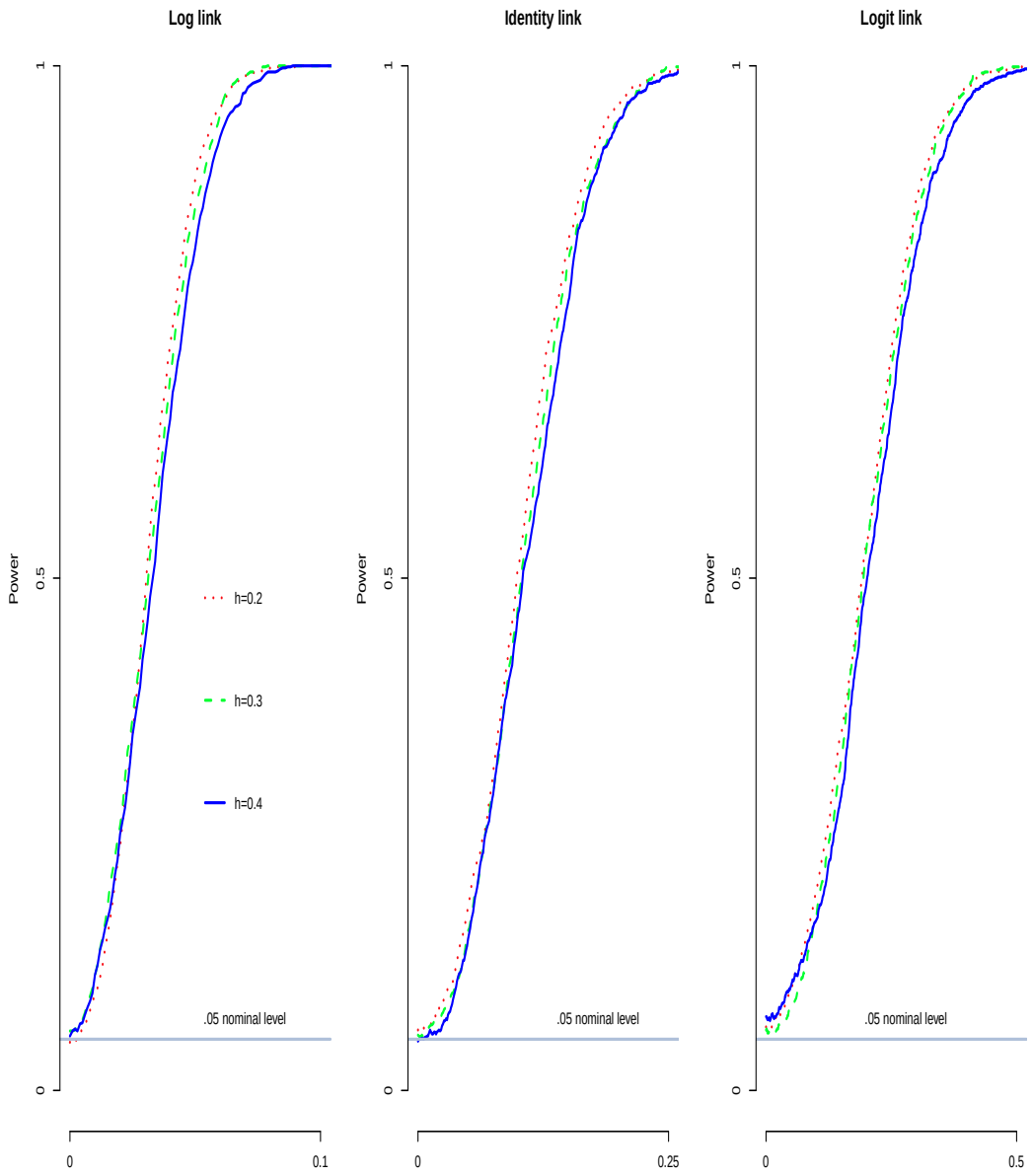


Figure 5: The power curves of the test for testing $\theta_0 = 0$ against $\theta_2 \neq 0$ with n=400 for log link function, identity link function and the logit link function, based on 500 simulations.

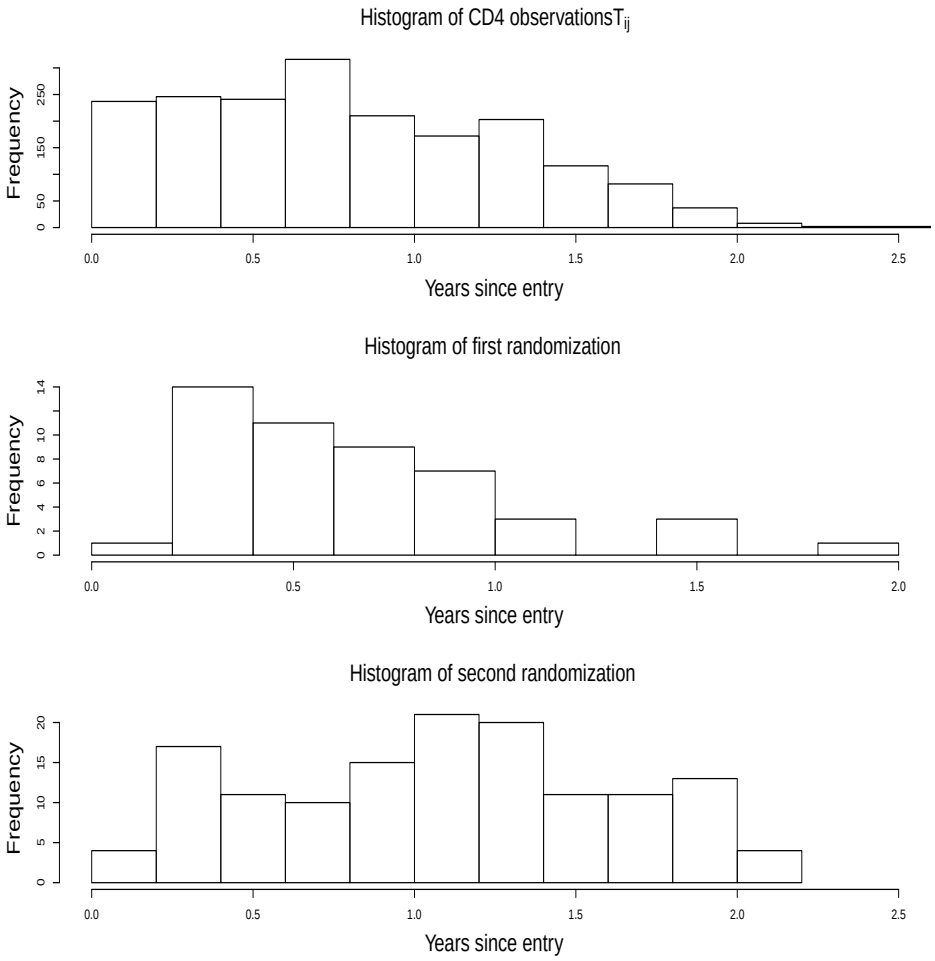


Figure 6: Histograms of time of visits, time of first randomization triggered by the codon 215 mutation, and time of second randomization triggered by the interim review while codon 215 wild-type.

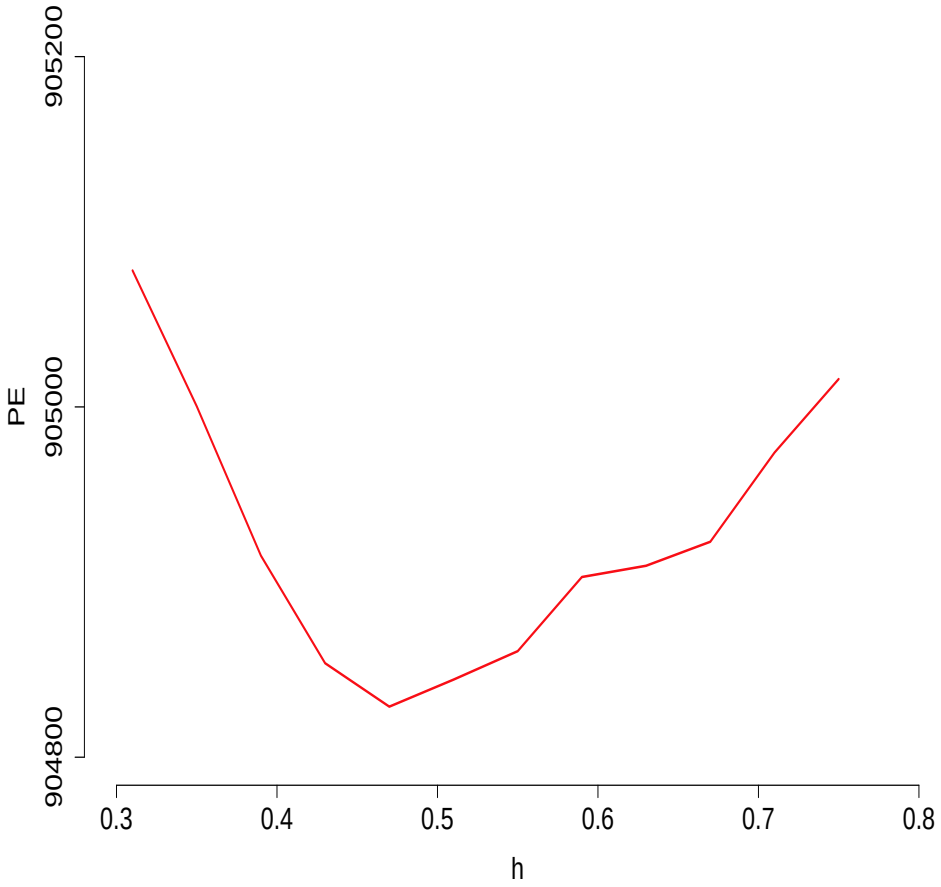


Figure 7: Prediction errors versus bandwidths, indicating the optimal bandwidth is around 0.47

Table 6: Demographics and Baseline Characteristics for ACTG 244 data.

		Number of subjects	Percentage
n		289	100 %
Sex			
	Male	246	85 %
	Female	43	14 %
Race/Ethnicity			
	White Non-Hispanic	181	62%
	Black Non-Hispanic	84	29 %
	Hispanic (Regardless of Race)	17	5 %
	Asian, Pacific Islander	5	1 %
	American Indian, Alaskan Native	2	0 %
Race			
	Black or African American	84	29 %
	White	182	62 %
	Unknown	23	7 %
Treatment Assignment			
	ZDV (1st Rand.)	17	6 %
	ZDV+DDI (1st Rand.)	15	5 %
	ZDV+DDI+NVP (1st Rand.)	17	6 %
	ZDV+DDI (2nd Rand.)	69	24 %
	ZDV+DDI+NVP (2nd Rand.)	68	24 %

Table 7: Estimated effects of switching treatments after drug-resistant virus was detected based on the ACTG 244 data. Point and 95% confidence interval estimates of β_1, β_2, β_3, β_4, θ_1, θ_2, θ_3, θ_4, θ_5 and θ_6 for model (2.16) based on the ACTG 244 data using $h = 0.47$ and unit weight.

	Estimate	Standard deviation	95% Confidence limits		p-value
β_1	-1.3243	0.7181	-2.7318	0.0832	0.0652
β_2	0.0622	0.0285	0.0064	0.1180	0.0289
β_3	-0.7315	0.7838	-2.2678	0.8048	0.3507
β_4	-0.7611	0.8607	-2.4482	0.9259	0.3765
θ_1	0.9899	1.2725	-1.5042	3.4841	0.4366
θ_2	-3.0354	1.9689	-6.8944	0.8236	0.1231
θ_3	-0.6186	0.9281	-2.4377	1.2005	0.5051
θ_4	-0.4806	1.3129	-3.0538	2.0926	0.7143
θ_5	-1.0576	0.9562	-2.9318	0.8167	0.2687
θ_6	0.3953	0.7520	-1.0787	1.8693	0.5992

Table 8: Estimated effects of switching treatments after drug-resistant virus was detected based on the ACTG 244 data. Point and 95% confidence interval estimates of β_1, β_2, β_3, β_4, θ_1, θ_2, θ_3, θ_4, θ_5 and θ_6 for model (2.16) based on the ACTG 244 data using $h = 0.47$ and calculated weight.

	Estimate	Standard deviation	95% Confidence limits		p-value
β_1	-1.2582	0.6872	-2.6051	0.0888	0.0671
β_2	0.0507	0.0254	0.0008	0.1006	0.0464
β_3	-0.6630	0.7362	-2.1060	0.7800	0.3679
β_4	-0.7310	0.8192	-2.3367	0.8747	0.3722
θ_1	0.3639	1.2144	-2.0162	2.7441	0.7644
θ_2	-2.7324	1.8128	-6.2856	0.8207	0.1317
θ_3	-0.3989	0.7709	-1.9100	1.1122	0.6049
θ_4	-0.6121	1.2672	-3.0958	1.8715	0.6290
θ_5	-1.1150	0.9232	-2.9245	0.6944	0.2271
θ_6	0.7096	0.7589	-0.7779	2.1971	0.3498

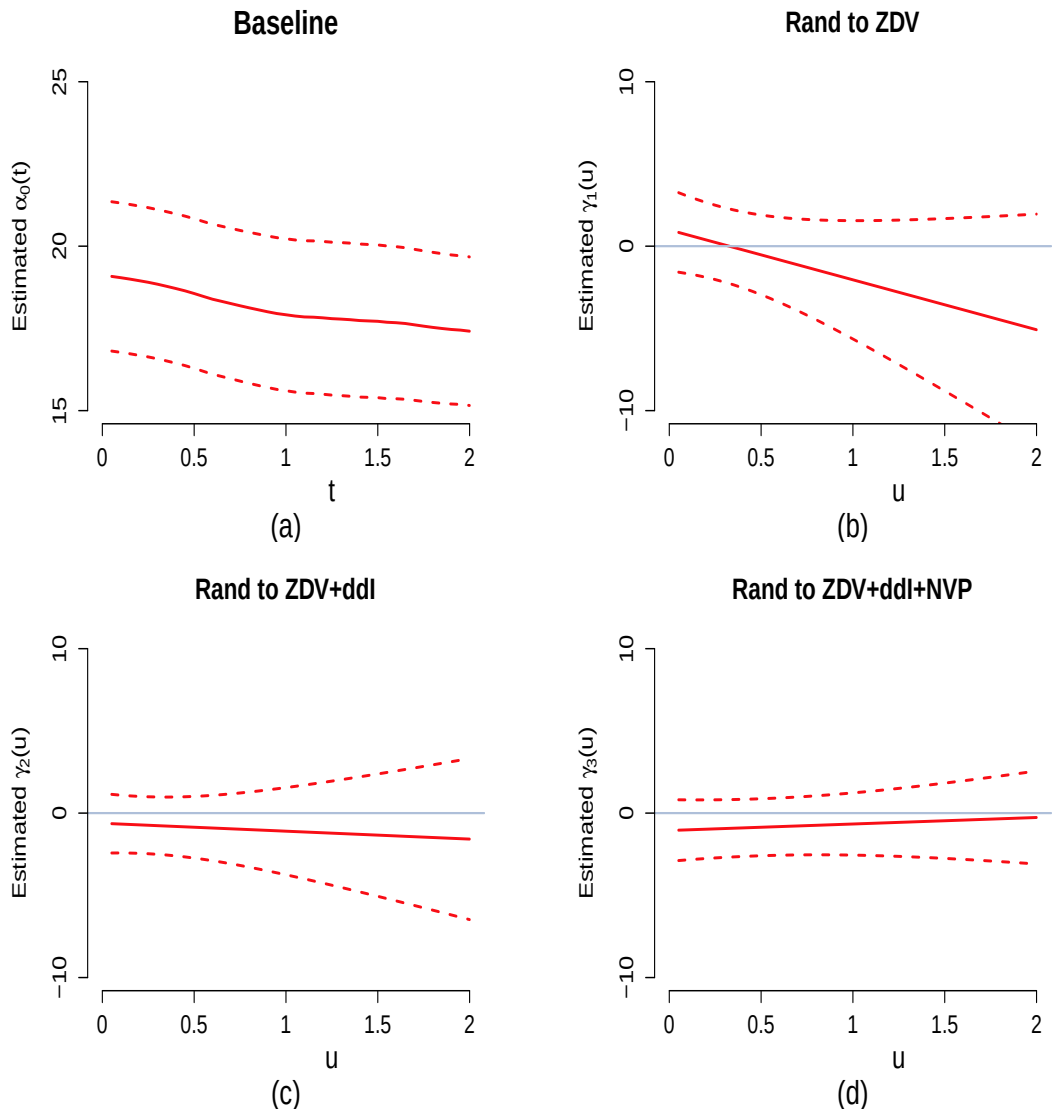


Figure 8: Estimated effects of switching treatments after drug-resistant virus was detected based on the ACTG 244 data. (a) is the estimated baseline function $\hat{\alpha}_0(t)$ with 95% pointwise confidence intervals; (b), (c) and (d) are the point and 95% confidence interval estimates of $\gamma_k(u)$, $k = 1, 3, 5$, respectively, under model (2.16) using $h = 0.47$.

Table 9: Estimated effects of switching treatments before drug-resistant virus was detected based on the ACTG 244 data. Point and 95% confidence interval estimates of β_1, β_2, β_3, β_4, θ_1, θ_2, θ_3, θ_4, θ_5 and θ_6 for model (2.17) based on the ACTG 244 data using $h = 0.47$ and unit weight.

	Estimate	Standard deviation	95% Confidence limits		p-value
β_1	-0.6660	0.6477	-1.9355	0.6036	0.3039
β_2	0.0198	0.0251	-0.0293	0.0689	0.4287
β_3	-1.0917	0.7225	-2.5079	0.3245	0.1308
β_4	-1.7661	0.7765	-3.2880	-0.2441	0.0229
θ_1	0.2564	0.5387	-0.7994	1.3123	0.6340
θ_2	3.1732	1.0324	1.1497	5.1967	0.0021
θ_3	1.0623	0.4257	0.2280	1.8967	0.0126
θ_4	2.7819	1.0759	0.6732	4.8906	0.0097

Table 10: Estimated effects of switching treatments before drug-resistant virus was detected based on the ACTG 244 data. Point and 95% confidence interval estimates of β_1, β_2, β_3, β_4, θ_1, θ_2, θ_3, θ_4, θ_5 and θ_6 for model (2.17) based on the ACTG 244 data using $h = 0.47$ and calculated weight.

	Estimate	Standard deviation	95% Confidence limits		p-value
β_1	-0.7090	0.6510	-1.9849	0.5670	0.2761
β_2	0.0194	0.0239	-0.0275	0.0662	0.4176
β_3	-1.0340	0.6924	-2.3912	0.3231	0.1353
β_4	-1.6997	0.7501	-3.1698	-0.2295	0.0235
θ_1	0.1065	0.4998	-0.8732	1.0861	0.8313
θ_2	3.3040	1.0166	1.3116	5.2965	0.0012
θ_3	0.8839	0.4070	0.0861	1.6817	0.0299
θ_4	2.7652	1.0956	0.6178	4.9126	0.0116

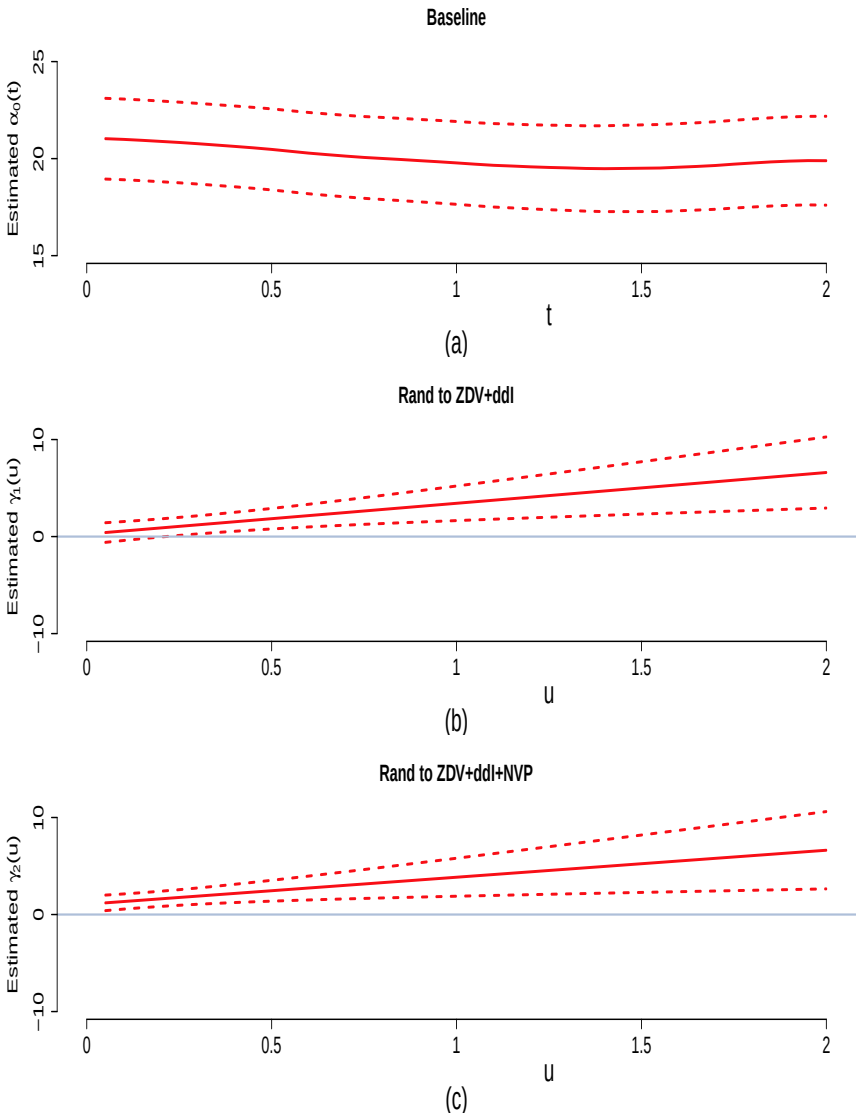


Figure 9: Estimated effects of switching treatments before drug-resistant virus was detected based on the ACTG 244 data. (a) is the estimated baseline function $\hat{\alpha}_0(t)$ with 95% pointwise confidence intervals; (b) and (c) are the point and 95% confidence interval estimates of $\gamma_k(u)$, $k = 1, 3$, respectively, under model (2.17) using $h = 0.47$.

CHAPTER 3: SEMIPARAMETRIC MODEL WITH NONPARAMETRIC COVARIATE-VARYING EFFECTS

Parametric modeling of covariate-varying effects reduces the infinite dimensional unknown parameters to a finite number of unknown parameters while permitting evaluation of the effects of switching treatments. The methods are useful in evaluating the effects of treatment switching when the number of patients switched treatment is not large enough for nonparametric estimation. However, nonparametric modeling of covariate-varying effects would provide greater flexibility.

3.1 Model

Suppose that there is a random sample of n subjects and τ is the end of follow-up. Suppose that observations of response process $Y_i(t)$ for subject i are taken at the sampling time points $0 \leq T_{i1} < T_{i2} < \cdots < T_{in_i} \leq \tau$, where n_i is the total number of observations on subject i. The sampling times are often irregular and depend on covariates. In addition, some subjects may drop out of the study early. Let $N_i(t) = \sum_{j=1}^{n_i} I(T_{ij} \leq t)$ be the number of observations taken on the ith subject by time t, where $I(\cdot)$ is the indicator function. Let C_i be the end of follow-up time or censoring time whichever comes first. The responses for subject i can only be observed at the time points before C_i. Thus $N_i(t)$ can be written as $N_i^*(t \wedge C_i)$, where $N_i^*(t)$ is the counting process of sampling times. Let $X_i(t)$ and $U_i(t)$ be the possibly time-dependent covariates associated with the ith subject. Suppose $U_i(t)$

has support $\mathcal{U}$. Assume that $\{Y_i(\cdot), X_i(\cdot), U_i(\cdot), N_i(\cdot), i = 1, \cdots, n\}$ are independent identically distributed (iid) random processes. The censoring time C_i is noninformative in the sense that $E\{dN_i^*(t) \mid X_i(t),\, U_i(t), C_i \geq t\} = E\{dN_i^*(t) | X_i(t), U_i(t)\}$ and $E\{Y_i(t)|X_i(t), U_i(t), C_i \geq t\} = E\{Y_i(t)|X_i(t),\, U_i(t)\}$. Assume that $dN_i^*(t)$ is independent of $Y_i(t)$ conditional on $X_i(t)$, $U_i(t)$ and $C_i \geq t$. The censoring time C_i is allowed to depend on $X_i(\cdot)$ and $U_i(\cdot)$.

Suppose that $X_i(t) = (X_{1i}^T(t), X_{2i}^T(t), X_{3i}^T(t))^T$ consist of three parts of dimensions p_1, p_2 and p_3, respectively, over the time interval $[0, \tau]$. Let $U_i(t)$ be the scalar covariate process. We propose the following generalized semiparametric regression model with varying coefficients:

$$\mu_i(t) = E\{Y_i(t)|X_i(t), U(t)\} = g^{-1}\{\alpha^T(t)X_{1i}(t) + \beta^T X_{2i}(t) + \gamma^T(U_i(t))X_{3i}(t)\}, \quad (3.1)$$

for $0 \leq t \leq \tau$, where $g(\cdot)$ is a known link function, $\alpha(\cdot)$ is a p_1-dimensional vector of completely unspecified functions, β is a p_2-dimensional vector of unknown parameters and $\gamma(U_i(t))$ is a p_3-dimensional vector of functions. The notation θ^T represents transpose of a vector or matrix θ. The first component of $X(t)$ is set to be 1, which gives a nonparametric baseline function. $\gamma(u)$ represents the effect of $X_{3i}(t)$ at the level u of a confounding covariate $U_i(t)$. Including the first component of $X_{3i}(t)$ also allows nonparametric modeling of covariate $U_i(t)$.

3.2 Estimation

Assume that $\alpha(\cdot)$ and $\gamma(\cdot)$ are smooth so that their first and second derivatives $\dot{\alpha}(\cdot)$, $\dot{\gamma}(\cdot)$, $\ddot{\alpha}(\cdot)$ and $\ddot{\gamma}(\cdot)$ exist. By the first order Taylor expansion, for $t \in \mathcal{N}_{t_0}$, a

neighborhood of t_0,

$$\alpha(t) = \alpha(t_0) + \dot{\alpha}(t_0)(t - t_0) + O((t - t_0)^2).$$

The first order Taylor expansion of $\gamma(u)$ in the neighborhood of u_0 denoted by $\mathcal{N}_{u_0}$ yields the approximation

$$\gamma(u) = \gamma(u_0) + \dot{\gamma}(u_0)(u - u_0) + O((u - u_0)^2).$$

For $t \in \mathcal{N}_{t_0}$ and $U_i(t) \in \mathcal{N}_{u_0}$, model (3.1) can be approximated by

$$\tilde{\mu}(t, t_0, u_0, \vartheta^*(t_0, u_0)|X_i, U_i) = \varphi\{\vartheta^{*T}(t_0, u_0)\tilde{X}_i^*(t, t_0, u_0) + \beta X_{2i}(t)\},$$

where $\vartheta^*(t_0, u_0) = (\alpha^T(t_0), \gamma^T(u_0), \dot{\alpha}^T(t_0), \dot{\gamma}^T(u_0))^T$, $\tilde{X}_i^*(t, t_0) = (X_{1i}^T(t), X_{3i}^T(t), X_{1i}^T(t) \times (t - t_0), X_{3i}^T(t) \times (U_i(t) - u_0))^T$.

At each t_0 and u_0, and for fixed β, we consider the following estimating function

$$U_\vartheta(\vartheta^*; t_0, u_0, \beta) = \sum_{i=1}^{n} \int_0^\tau W_i(t) \left[Y_i(t) - \tilde{\mu}(t, t_0, u_0, \vartheta^*(t_0, u_0)|X_i, U_i)\right] \times \tilde{X}_i^*(t, t_0, u_0) K_{h,b}(t, U_i(t); t_0, u_0)\, dN_i(t), \qquad (3.2)$$

where $K_{h,b}(t, U_i(t); t_0, u_0) = K_h(t - t_0)K_b(U_i(t) - u_0)$ is a two dimensional product kernel with bandwidths h and b, $K_h(\cdot) = K_1(\cdot/h)/h$ and $K_b(\cdot) = K_2(\cdot/b)/b$, $K_1(\cdot)$ and $K_2(\cdot)$ are kernel functions, $h = h_n$ and $b = b_n$ are bandwidth parameters. The solution to $U(\vartheta^*, t_0, u_0, \beta) = 0$ is denoted by $\tilde{\vartheta}^*(t_0, u_0, \beta)$.

Let $\tilde{\vartheta}(t, u, \beta)$ be the first $p_1 + p_3$ components of the solution of (3.2) given β. The profile least-squares estimator $\hat{\beta}$ is obtained by minimizing the following profile least-

squares function:

$$\ell_\beta(\beta) = \sum_{i=1}^{n} \int_{t_1}^{t_2} Q_i(t) \left[Y_i(t) - \varphi\{\tilde{\vartheta}(t, U_i(t), \beta)\tilde{X}_i(t) + \beta^T X_{2i}(t)\} \right]^2 dN_i(t), \qquad (3.3)$$

where $Q_i(t)$ is a nonnegative weight process and $\tilde{X}_i(t) = ((X_{1i}(t))^T, (X_{3i}(t))^T)^T$. The Newton-Raphson iterative method can be used to find the estimator of β. Taking derivative with respect to β, we obtain the profile estimating equation for β:

$$\begin{aligned} U_\beta(\beta) = \sum_{i=1}^{n} \int_{t_1}^{t_2} & W_i(t) \left[Y_i(t) - \varphi\{\tilde{\vartheta}(t, U_i(t), \beta)\tilde{X}_i(t) + \beta^T X_{2i}(t)\} \right] \\ & \times \left\{ \frac{\partial \tilde{\vartheta}(t, U_i(t), \beta)}{\partial \beta} \tilde{X}_i(t) + X_{2i}(t) \right\} dN_i(t), \end{aligned} \qquad (3.4)$$

where $\frac{\partial \tilde{\vartheta}(t, U_i(t), \beta)}{\partial \beta}$ is the first $p_1 + p_3$ rows of

$$\frac{\partial \tilde{\vartheta}^*(t, U_i(t), \beta)}{\partial \beta} = - \left\{ \frac{\partial U_\vartheta(\vartheta^*; t, U_i(t), \beta)}{\partial \vartheta^*} \right\}^{-1} \frac{\partial U_\vartheta(\vartheta^*; t, U_i(t), \beta)}{\partial \beta} \Bigg|_{\vartheta^* = \tilde{\vartheta}^*(t, U_i(t), \beta)}.$$

The estimators $\hat{\vartheta}^*(t_0, u_0)$ can be obtained through an iterated estimation procedure. Let $\hat{\alpha}(t_0, u_0)$ include the first p_1 elements of $\hat{\vartheta}^*(t_0, u_0)$ corresponding to the position of $\alpha(t_0)$ in $\vartheta^*(t_0, u_0)$. Let $\hat{\gamma}(t_0, u_0)$ include the elements of of $\hat{\vartheta}^*(t_0, u_0)$ corresponding to the position of $\gamma(u_0)$ in $\vartheta^*(t_0, u_0)$. Estimation of $\alpha(t_0)$ by $\hat{\alpha}^*(t_0, u_0)$ is inefficient because it only utilizes the local observations with $U_i(t) \in \mathcal{N}_{u_0}$ for $t \in \mathcal{N}_{t_0}$. More efficient estimator for $\alpha(t_0)$ at t_0 can be obtained through aggregation without restricting $U_i(t) \in \mathcal{N}_{u_0}$. Similarly, more efficient estimator of $\gamma(u_0)$ can be obtained through aggregating over t such that $U_i(t) = u_0$. We propose the following estimators $\hat{\alpha}(t_0)$ and $\hat{\gamma}(u_0)$ for $\alpha(t_0)$ and $\gamma(u_0)$, respectively:

$$\hat{\alpha}(t_0) = n^{-1} \sum_{j=1}^{n} \hat{\alpha}^*(t_0, U_j(t_0)), \qquad \hat{\gamma}(u_0) = n_{u_0}^{-1} \sum_{j=1}^{n_{u_0}} \hat{\gamma}^*(t_{u_0,j}, u_0), \qquad (3.5)$$

where $t_{u_0,j} \in U_j^{-1}(u_0) = \{t : U_j(t) = u_0\}$, and n_{u_0} is the number of points in the union $\cup_{j=1}^{n}\{U_j^{-1}(u_0)\}$. For the first motivating example, $U_j(t) = t - S_j$, thus $t_{u_0,j} = u_0 + S_j$. If it is difficult to find $U_j^{-1}(u_0)$, $\gamma(u_0)$ can also be estimated by

$$\hat{\gamma}(u_0) = \int_0^{\tau} \hat{\gamma}^*(t, u_0)\, dt. \tag{3.6}$$

Computational algorithm

The estimators $\hat{\alpha}(t_0)$, $\hat{\gamma}(u_0)$ and $\hat{\zeta}$ can be accomplished through the following iterated algorithm:

1. Given $\hat{\vartheta}(t,u)^{\{0\}}$ and $\hat{\beta}^{\{0\}}$ as the initial values;

2. For each jump point of $\{N_i(\cdot), i = 1, \cdots, n\}$, say t and $u = U_i(t)$, the mth step estimator $\hat{\vartheta}^{*\{m\}}(t,u) = \hat{\vartheta}^*(t,u,\hat{\beta}^{\{m-1\}})$ is the root of the estimating function (3.2) satisfying $U_\vartheta(\hat{\vartheta}^{*\{m\}}(t,u), t, u, \hat{\beta}^{\{m-1\}}) = 0$, where $\hat{\beta}^{\{m-1\}}$ be the estimate of β at the $(m-1)$th step.

3. The mth step estimator $\hat{\beta}^{\{m\}}$ is minimizer of (3.3) obtained after replacing $\hat{\vartheta}(t,u)$ with $\hat{\vartheta}^{\{m\}}(t,u)$.

4. Repeating step 2 and 3, the estimators $\hat{\vartheta}^{*\{m\}}(t,u)$ and $\hat{\beta}^{\{m\}}$ are updated at each iteration until converges. $\hat{\beta}$ is $\hat{\beta}^{\{m\}}$ at the convergence.

5. The estimate of $\hat{\alpha}(t_0)$ and $\hat{\gamma}(u_0)$ is obtained by (3.5) at the grid points t_0 and u_0 fine enough such that their plots look reasonably smooth.

3.3 Asymptotics

3.3.1 Notations

Let $\mathscr{I}_1 = \{\mathscr{I}_{jk}\}$ be a $p_1 \times (p_1 + p_3)$ matrix with elements $\mathscr{I}_{jk} = 1$ for $j = 1, \ldots, p_1$, $k = j$, and $\mathscr{I}_{jk} = 0$ otherwise. And let $\mathscr{I}_3 = \{\mathscr{I}_{jk}\}$ be a $p_3 \times (p_1 + p_3)$ matrix with elements $\mathscr{I}_{jk} = 1$ for $j = 1, \ldots, p_3$, $k = j + p_1$, and $\mathscr{I}_{jk} = 0$ otherwise.

Let $\alpha_0(t)$, β_0 and $\gamma_0(u)$ be the true value of $\alpha(t)$, β and $\gamma(u)$ under model (3.1), respectively. Let $\mu_i(t) = \varphi\{\alpha_0(t)X_{1i}(t) + \beta_0 X_{2i}(t) + \gamma_0(U_i(t))X_{3i}(t)\}$ and $\dot{\mu}_i(t) = \dot{\varphi}\{\alpha_0(t)X_{1i}(t) + \beta_0 X_{2i}(t) + \gamma_0(U_i(t))X_{3i}(t)\}$. Define

$$e_{11}(t,u) = E[\omega_i(t)\dot{\mu}_i(t)\{\tilde{X}_i(t)\}^{\otimes 2}\xi_i(t)\lambda_i(t)|U_i(t) = u]f_U(t,u)$$

and

$$e_{12}(t,u) = E[\omega_i(t)\dot{\mu}_i(t)\tilde{X}_i(t)(X_{2i}(t))^T\xi_i(t)\lambda_i(t))|U_i(t) = u]f_U(t,u),$$

where $\xi_i(t) = I(C_i \geq t)$ and $f_U(t,u)$ is the density function of $U(t)$ evaluated at u.

Let $\hat{\mu}_i(t) = \varphi\{\hat{\vartheta}^T(t, U_i(t))\tilde{X}_i(t) + \beta^T X_{2i}(t)\}$ and $\hat{\dot{\mu}}_i(t) = \dot{\varphi}\{\hat{\vartheta}^T(t, U_i(t))\tilde{X}_i(t) + \beta^T X_{2i}(t)\}$. Let

$$\hat{E}_{11}(t_0,u_0) = n^{-1}\sum_{i=1}^{n}\int_0^\tau K_h(t-t_0)K_b(U_i(t)-u_0)W_i(t)\hat{\dot{\mu}}_i(t)\{\tilde{X}_i(t)\}^{\otimes 2}\,dN_i(t),$$

and

$$\hat{E}_{12}(t_0,u_0) = n^{-1}\sum_{i=1}^{n}\int_0^\tau K_h(t-t_0)K_b(U_i(t)-u_0)W_i(t)\hat{\dot{\mu}}_i(t)\tilde{X}_i(t)(X_{2i}(t))^T\,dN_i(t).$$

3.3.2 Asymptotic Properties

The following theorems present the asymptotic properties of the proposed estimators.

Theorem 3.1. *Assume that Condition II holds. Then* $\sqrt{n}(\hat{\beta} - \beta_0)$ *converges in distribution to a mean-zero normal distribution* $N(0, A_\beta^{-1}\Sigma_\beta A_\beta^{-1})$*, where*

$$A_\beta = E\left[\int_0^\tau \omega_i(t)\dot{\mu}_i(t)\{X_{2i}(t) - (e_{12}(t, U_i(t)))^T(e_{11}(t, U_i(t)))^{-1}\tilde{X}_i(t)\}^{\otimes 2}\, dN_i(t)\right]$$

and

$$\Sigma_\beta = E\left(\int_{t_1}^{t_2} \omega_i(t)[Y_i(t) - \mu_i(t)]\{X_{2i}(t) - (e_{12}(t, U_i(t)))^T(e_{11}(t, U_i(t)))^{-1}\tilde{X}_i(t)\}\, dN_i(t)\right)^{\otimes 2}.$$

The matrix A_β can be consistently estimated by

$$\hat{A}_\beta = n^{-1}\sum_{i=1}^{n}\int_{t_1}^{t_2} W_i(t)\hat{\dot{\mu}}_i(t)\{X_{2i}(t) - (\hat{E}_{12}(t, U_i(t)))^T(\hat{E}_{11}(t, U_i(t)))^{-1}\tilde{X}_i(t)\}^{\otimes 2}\, dN_i(t)$$

and Σ_β can be consistently estimated by

$$\begin{aligned}\hat{\Sigma}_\beta = n^{-1}\sum_{i=1}^{n}\Big(\int_{t_1}^{t_2} & W_i(t)\{Y_i(t) - \hat{\mu}_i(t)\} \\ & \times \{X_{2i}(t) - (\hat{E}_{12}(t, U_i(t)))^T(\hat{E}_{11}(t, U_i(t)))^{-1}\tilde{X}_i(t)\}\, dN_i(t)\Big)^{\otimes 2}. \qquad (3.7)\end{aligned}$$

Theorem 3.2. *Assume that Condition II holds. Then*

(a) $\sup_{t\in[0,\tau]} |\hat{\alpha}(t) - \alpha_0(t)| = o_p(1)$*;*

(b) $\sqrt{nh}(\hat{\alpha}(t) - \alpha_0(t) - \frac{1}{2}h^2\nu_2\ddot{\alpha}(t)) \xrightarrow{\mathcal{D}} N\left(0, \Sigma_\alpha(t)\right),$

where

$$\Sigma_\alpha(t) = \lim_{n\to\infty} hE\left[\int_0^\tau \omega_i(t)\left\{Y_i(s) - \mu_i(s)\right\}\mathscr{I}_1 e_{11}(t, U_i(s))^{-1} X_i(s) K_h(s-t)\, dN_i(s)\right]^{\otimes 2}.$$

The limiting variance-covariance matrix $\Sigma_\alpha(t)$ can be consistently estimated by

$$\frac{h}{n}\sum_{i=1}^{n}\left[\int_0^\tau W_i(t)\left\{Y_i(s) - \hat{\mu}_i(s)\right\}\mathscr{I}_1 \hat{E}(t, U_i(s))^{-1} X_i(s) K_h(s-t)\, dN_i(s)\right]^{\otimes 2}.$$

Theorem 3.3. *Assume that Condition II holds. Then*

(a) $\sup_{u\in[u_1,u_2]} |\hat{\gamma}(u) - \gamma_0(u)| = o_p(1)$;

(b) $\sqrt{nb}(\hat{\gamma}(u) - \gamma_0(u) - \frac{1}{2}b^2\nu_2\ddot{\gamma}(u)) \xrightarrow{\mathcal{D}} N\left(0, \Sigma_\gamma(u)\right).$

where

$$\Sigma_\gamma(u) = \lim_{n\to\infty} bE\left[\int_0^\tau \omega_i(t)\left\{Y_i(t) - \mu_i(t)\right\}\mathscr{I}_3 e_{11}(t, u)^{-1} X_i(t) K_b(U_i(t) - u)\, dN_i(t)\right]^{\otimes 2}.$$

The limiting variance-covariance matrix $\Sigma_\gamma(u)$ can be consistently estimated by

$$\frac{b}{n}\sum_{i=1}^{n}\left[\int_0^\tau W_i(t)\left\{Y_i(t) - \hat{\mu}_i(t)\right\}\mathscr{I}_3 \hat{E}(t, u)^{-1} X_i(t) K_b(U_i(t) - u)\, dN_i(t)\right]^{\otimes 2}.$$

Let $\Gamma_0(u) = \int_{u_1}^{u} \gamma_0(s)ds$ and $\hat{\Gamma}(u) = \int_{u_1}^{u} \hat{\gamma}(s)ds$. The following theorem presents a weak convergence for $G_n(u) = n^{1/2}(\hat{\Gamma}(u) - \Gamma_0(u))$ over $u \in [u_1, u_2] \subset \mathcal{U}$.

Theorem 3.4. *Assume that Condition II holds, we have* $G_n(u) = n^{-1/2}\sum_{i=1}^{n} H_i(u) +$

$o_p(1)$ *uniformly in* $u \in [u_1, u_2]$, *where*

$$H_i(u) = \int_{u_1}^{u} \int_0^{\tau} \omega_i(t) \{Y_i(t) - \mu_i(t)\} \mathscr{I}_3 e_{11}(t,s)^{-1} X_i(t) K_b(U_i(t) - s)\, dN_i(t)\, ds$$
$$+ \int_{u_1}^{u} E\{(e_{11}(U_j^{-1}(s), s))^{-1} e_{12}(U_j^{-1}(s), s)\}\, ds A_\beta^{-1} \int_{t_1}^{t_2} \omega_i(t)[Y_i(t) - \mu_i(t)]$$
$$\times \{X_{2i}(t) - (e_{12}(t, U_i(t)))^T (e_{11}(t, U_i(t)))^{-1} \tilde{X}_i(t)\}\, dN_i(t)$$

The processes $G_n(u)$ *converges weakly to a zero-mean Gaussian process on* $[u_1, u_2]$.

3.3.3 Hypothesis Testing of $\gamma(u)$

In this section we assume $X_3(t)$ is one dimensional for simplicity and let $\Gamma(u)$ be the cumulative coefficient function. A formal hypothesis testing procedure can be established to check whether there are varying effects of $X_3(t)$, that is to test

$H_0^{(1)}$: $\gamma(u) = 0$ for $u \in [u_1, u_2]$

$H_a^{(1)}$: $\gamma(u) \neq 0$ for some u.

In Theorem 3.4, we know that $G_n(u)$, $u \in [u_1, u_2]$ converges weakly to a mean zero Gaussian process with continuous sample paths on $u \in [u_1, u_2]$. Further, the distribution of $G_n(u)$, for $u \in [u_1, u_2]$, can be approximated by using the Gaussian multipliers resampling method based on $G^{*(1)}(u) = n^{-1/2} \sum_{i=1}^{n} \hat{H}_i(u)\phi_i$, where $\phi_1, \cdots, \phi_n$ are repeatedly generated independent normal random variables and $\hat{H}_i(t)$ are obtained by replacing the unknown quantities in $H_i(t)$ with their corresponding empirical counterparts.

Consider the test process $Q^{(1)}(u) = n^{1/2}\hat{\Gamma}(u)$, $u \in [u_1, u_2]$. Then $Q^{(1)}(u) = G_n(u) + n^{1/2}\Gamma_0(u)$, $u \in [u_1, u_2]$. Under H_0, $\Gamma_0(u) = 0$ for $u \in [u_1, u_2]$, which motivates the

following supreme type and integrated squared difference type test statistics:

$$S_1^{(1)} = \sup_{u\in[u_1,u_2]} |Q^{(1)}(u)|,$$

and

$$S_2^{(1)} = \int_{[u_1,u_2]} Q^{(1)}(u)^2\, du.$$

Under H_0, the distribution of $Q^{(1)}(u)$, $u \in [u_1, u_2]$, can be approximated by the conditional distribution of $G^*(u)$, $u \in [u_1, u_2]$, given the observed data sequence. Hence, the distributions of $S_1^{(1)}$ and $S_2^{(1)}$ under $H_0^{(1)}$ can be approximated by the conditional distributions of

$$S_1^{*(1)} = \sup_{u\in[u_1,u_2]} |G^*(u)|,$$

and

$$S_2^{*(1)} = \int_{[u_1,u_2]} G^*(u)^2\, du,$$

given the observed data sequence, respectively. The critical values c_1 and c_2 of the test statistics $S_1^{(1)}$ and $S_2^{(1)}$ can be approximated by the $(1-\alpha)$-quantile of $S_1^{*(1)}$ and $S_2^{*(1)}$, which can be obtained by repeatedly generating a large number, say 500, of independent sets of normal samples $\{\phi_i,\, i = 1 \cdots, n\}$ while holding the observed data sequence fixed. At significant level α, the tests based on $S_1^{(1)}$ and $S_2^{(1)}$ reject H_0 if $S_1^{(1)} > c_1^{(1)}$ and $S_2^{(1)} > c_2^{(1)}$ respectively.

Nonparametric estimation is flexible but it yeilds to slower convergent rate. One may interested in parametric estimation. One important usage of the nonparametric estimation is to verify the functional form used in parametric estimation.

Let $\hat{\Gamma}(t) = \int_{t_1}^{t} \hat{\gamma}(u)du$ and its parametric counterpart $\Gamma(t;\hat{\theta}) = \int_{t_1}^{t} \gamma(u;\hat{\theta})du$. A large deviation between $\hat{\Gamma}(u)$ and $\Gamma(u;\hat{\theta})$ would indicate the lack of fit of the lack of fit of the parametric form $\Gamma(u;\hat{\theta})$.

Let $Q^{(2)}(u) = n^{1/2}(\hat{\Gamma}(u) - \Gamma(u;\hat{\theta}))$ To test the null hypothesis

$H_0^{(2)}$: $\gamma(u) = \gamma(u;\theta)$ for $u \in [u_1, u_2]$

$H_a^{(2)}$: $\gamma(u) \neq \gamma(u;\theta)$ for some u.

we consider the following supremum type and integrated squared difference type test statistics:

$$S_1^{(2)} = \sup_{u\in[u_1,u_2]} |Q^{(2)}(u)|,$$

and

$$S_2^{(2)} = \int_{[u_1,u_2]} Q^{(2)}(u)^2\, dt,$$

Under these tests, the coefficient regression functions $\gamma(u)$ can be tested simultaneously. By Theorem 3.4 we have

$$Q^{(2)}(u) = n^{-1/2} \sum_{i=1}^{n} H_i^{(2)}(u) + o_p(1), \tag{3.8}$$

where

$$\begin{aligned} H_i^{(2)}(u) =& H_i(u) - \int_{u_1}^{u} \gamma_\theta'(s;\theta)\, ds A^{-1} \\ & \times \int_{t_1}^{t_2} \epsilon_i(t) \left\{ \frac{\partial \widetilde{\alpha}(t,\vartheta_0)}{\partial \vartheta} X_{1i}(t) + \frac{\partial \eta(U_i(t),\vartheta_0)}{\partial \vartheta} X_{2i}^*(t) \right\} dN_i(t). \end{aligned} \tag{3.9}$$

inheriting notations in Chapter 2. We obtain $\hat{H}_i^{(2)}(u)$ by replacing the unknown quantities in $H_i^{(2)}(u)$ with their corresponding empirical counterparts. The distribution of $Q^{(2)}(u)$ can be estimated by the conditional distribution of $G^{*(2)}(u) =$

$n^{-1/2}\sum_{i=1}^{n}\hat{H}_i^{(2)}(u)\phi_i$, where $\phi_1,\cdots,\phi_n$ are repeatedly generated independent standard normal random variables. The critical values of the test statistics $S_1^{(2)}$ and $S_2^{(2)}$ at the significant level α can be estimated by the $(1-\alpha)$-quantile of , say 500, copies of $S_1^{*(2)}$ and $S_2^{*(2)}$, respectively, where

$$S_1^{*(2)} = \sup_{u\in[u_1,u_2]} |G^{*(2)}(u)|,$$

and

$$S_2^{*(2)} = \int_{[u_1,u_2]} G^{*(2)}(u)^2\, du.$$

The null hypothesis is rejected if the test statistics are greater than the critical values.

3.3.4 Bandwidth Selection

Instead of selecting the bandwidths by the leave-one-out cross-validation method suggested in Rice and Silverman (1991), we choose the bandwidths for the mean function estimator via a K-fold cross validation procedure (Tian et al., 2005) to reduce the computational cost. Below, we describe the K-fold cross-validation method for the bandwidth selection for two bandwidth parameters h and b. Supposing that subjects are randomly divided into K groups, $(D_1, D_2, \cdots, D_K)$, the K-fold cross-validation bandwidth is $(h_{opt,K}, b_{opt,K}) = \arg\min_{h,b}\sum_{k=1}^{K} PE_k(h,b)$, where the kth prediction error is given by

$$PE_k(h,b) = \sum_{i\in D_k}\Big\{\int_0^{\tau} W_i(t)[Y_i(t) - \varphi\{\hat{\alpha}_{(-k)}^T(t)X_{1i}(t) + \hat{\beta}_{(-k)}^T X_{2i}(t) + \hat{\gamma}_{(-k)}^T(U_i(t))X_{3i}(t)\}]dN_i(t)\Big\}^2 \tag{3.10}$$

for $k = 1, \cdots, K$, where $\hat{\alpha}_{(-k)}$, $\hat{\beta}_{(-k)}$ and $\hat{\gamma}_{(-k)}$ are estimated using the data excluding subjects in D_k.

3.4 Simulations

We conducted extensive simulation studies to assess the finite sample performance of the proposed methods. The proposed methods are illustrated under three models with three popular link functions below:

$$\text{Identity link}: \ Y_i(t) = \alpha(t) + \beta Z_i + \gamma(U_i(t))X_i + \epsilon(t); \tag{3.11}$$

$$\text{Logarithm link}: \ Y_i(t) = \exp\{\alpha(t) + \beta Z_i + \gamma(U_i(t))X_i\} + \epsilon(t); \tag{3.12}$$

$$\text{Logitlink}: \ \text{logit}\{P(Y_i(t) = 1)\} = \alpha(t) + \beta Z_i + \gamma(U_i(t))X_i, \tag{3.13}$$

for $0 \leq t \leq \tau$ with $\tau = 3.5$, where $\alpha(t) = 0.5\sqrt{t}$, $\gamma(U_i(t)) = -0.6U_i(t)$, X_i is a uniform random variable on $[-1, 1]$, Z_i is a Bernoulli random variable with the success probability of 0.5, and $U_i(t) = t + S_i$ where S_i is a uniform random variable on $[0, 1]$. The error $\epsilon_i(t)$ has a normal distribution conditional on the ith subject with mean ϕ_i and variance 0.5^2, and ϕ_i is $N(0, 1)$. The observation time follows a Poisson process with the proportional mean rate model $h(t|X_i, S_i) = 2.5\exp(0.9Z_i)$. The censoring times C_i is generated from an uniform distribution on $[2.5, 8]$. There are approximately twelve observations per subject in $[0, \tau]$ and about 30% subjects are censored before $\tau = 3.5$. The Epanechnikov kernel $K(u) = .75(1 - u^2)I(|u| \leq 1)$ and the unit weight function are used. We take $t_1 = h/2$ and $t_2 = \tau - h/2$ in the estimating functions (3.3) to avoid larger variations on the boundaries. The suitable bandwidths are around $h = 0.45$ and $b = 0.475$ based on a preliminary investigation in which

the cross-validation method was applied to a few simulated data sets (see Figure 10). We report several bandwidths around $h = 0.45$ and $b = 0.475$ to investigate the sensitivity of the bandwidth selection.

The performances of $\hat{\beta}$, and the performances of $\hat{\alpha}(t)$ and $\hat{\gamma}(u)$ at a fixed point t and u are measured through the Bias, the sample standard error of the estimators (SEE), the sample mean of the estimated standard errors (ESE) and the 95% empirical coverage probability (CP). The overall performance of the estimator $\hat{\alpha}(t)$ are evaluated by the square root of integrated mean square error $RMSE_\alpha = \left\{ \frac{1}{N(\tau-2h)} \sum_{j=1}^{N} \int_h^{\tau-h} (\hat{\alpha}_j(t) - \alpha_0(t))^2 \, dt \right\}^{1/2}$, where N is the repetition number, $\hat{\alpha}_j(t)$ is the jth estimate of $\alpha(t)$ for $j = 1, \cdots, N$. $RMSE_\gamma$ is defined likewise.

Table 11-13 summarize the Bias, SEE, ESE and CP for β and RMSE for $\alpha(t)$ and $\gamma(t)$ under Model (3.11)-(3.13). Each entry of the tables is calculated based on 500 repetitions. It can be seen that the proposed estimations perform well for all three models. It appears that the estimates are unbiased and there is a good agreement between the estimated and empirical standard errors. Several bandwidths are selected and it shows that the results are not sensitive to the bandwidth selection. And in general smaller bandwidth leads to smaller variance but bigger bias. The bias and variances of the estimates decrease as the sample size increases. The CPs are close to the nominal value.

Figure 11-13 (a) and (b) show the bias of the time-varying coefficient estimates for $\alpha(t)$ and $\gamma(u)$ at selected time points. (c), (d) correspond to the SSEs and (e), (f) correspond to the ESEs. (g) and (h) correspond to the empirical CPs averaged over 500 simulations. We can see that the local estimators are close to the true values

and the ESE provides a good approximation for the SSE of the point estimates. The empirical coverage probabilities are reasonable and the results become better when the sample size increases.

It's of interest to test whether there are covariate-varying effects of $X_3(t)$ with respect of $U_i(t)$, which is to test $H_0 : \gamma(u) = 0$ for $u \in [u_1, u_2]$. The observed sizes of the test statistics are calculated under the null hypothesis. The powers of the test are calculated from $\theta = 0.01$ to 0.3 by 0.01 for $\gamma(U_i(t)) = -\theta U_i(t)$. A larger value of θ indicates an increased departure from the null hypothesis. The power curve of the test against θ at 5% nominal level is potted in Figure 14 with $n = 400$ for statistics $S_1^{(1)}$ and $S_2^{(1)}$.

3.5 Application to the ACTG 244 trial

We apply the methods developed in the previous sections to the randomized, double-blind AIDS Clinical Trials Group (ACTG) 244 trial to evaluate the clinical utility of monitoring for the ZDV^{R} mutation T215Y/F in HIV-1 reverse transcriptase in asymptomatic HIV-1-infected subjects taking ZDV monotherapy.

First, we examine the effects of switching treatments following detection of the T215Y/F mutation. The 3-fold cross-validation method for bandwidth selection yields h =0.47 Let Y be the square root of CD4 count, Z_1 be Sex (1 if Female; 0 if Male), Z_2 be Age in years, Z_3 and Z_4 be dummy variables coding race ($Z_3 = 1$ if white and 0 otherwise, $Z_4 = 1$ if black and 0 otherwise), S be the time of the codon 215 mutation, $Trt_{1i}(t) = 1$ if randomized to ZDV and 0 otherwise, $Trt_{2i}(t) = 1$ if randomized to ZDV+ddI and 0 otherwise, and $Trt_{3i}(t) = 1$ if randomized to ZDV+ddI+NVP and 0

otherwise; note that all three indicators are zero prior to detection of the mutation. After preliminary exploration of the data, we propose the following model for each subject i:

$$\begin{aligned} Y_i(t) =& \alpha_0(t) + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 Z_{4i} + \gamma_1(t - S_i)Trt_{1i}(t) \\ &+ \gamma_2(t - S_i)Trt_{2i}(t) + \gamma_3(t - S_i)Trt_{3i}(t) + \epsilon_i(t) \end{aligned} \tag{3.14}$$

for $t \in [0, 2]$. The 3-fold cross-validation method for bandwidth selection yields $h = 0.5$ and $b = 2.5$.

The estimated baseline function $\hat{\alpha}_0(t)$ and the time-varying switching-treatment effect function $\hat{\gamma}_k(u)$, $k = 1, 2, 3$ with their 95% pointwise confidence intervals are presented in Figure 15, where u is the time since T215Y/F mutation-based treatment switching in the first randomization. The time-invariant parameter estimates are presented in Table 14.

The results show that CD4 counts decrease over time. None of the constant effects are significant. The estimated $\hat{\gamma}_1(u)$ looks flat, the estimated $\hat{\gamma}_2(u)$ looks decreasing while the estimated $\hat{\gamma}_3(u)$ looks increasing. The hypothesis testing procedure developed in Section 3.3.3 is applied here to test whether $H_0 : \gamma_k(u) = 0$, $k = 1, 2, 3$ against $H_a : \gamma_k(u) > 0$, $k = 1, 2, 3$. The p-values using test statistics $S_1^{(1)}$ ($S_2^{(1)}$) are 0.6435 (0.6190), 0.9855 (0.9945) and 0.9670 (0.9225) for $k = 1, 2, 3$ respectively, which indicate to fail to reject the null hypothesis. It shows that none of the randomly assigned treatments increased CD4 counts significantly after the codon 215 mutation.

This analysis does not show the benefit of switching treatments (among those available in the study) after drug-resistant virus was detected. This result is consistent

with the result of Chapter 2.

Next, we examine the effects of switching treatments *before* drug-resistant virus, the codon 215 mutation, was detected. After independent review of the study data in September 1996, all subjects were offered randomization to the ZDV+ddI or ZDV+ddI+NVP arms with six months of additional follow-up.

We let S be the time of the second randomization after interim review and only include subjects without 215 mutation in the analysis. The model is similar to before:

$$\begin{aligned} Y_i(t) =& \alpha_0(t) + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 Z_{4i} \\ &+ \gamma_1(t - S_i) Trt_{1i}(t) + \gamma_2(t - S_i) Trt_{2i}(t) + \epsilon_i(t). \end{aligned} \tag{3.15}$$

The 3-fold cross-validation method for bandwidth yields $h = 0.5$ and $b = 1.5$. The results for constant effects are in Table 15. In addition, Figure 16 shows that for each treatment switch to ZDV+ddI or ZDV+ddI+NVP, CD4 cell counts significantly rise.

The p-values using test statistics $S_1^{(1)}$ ($S_2^{(1)}$) to test $H_0 : \gamma_k(u) = 0$, $k = 1, 2$ against $H_a : \gamma_k(u) \neq 0$, $k = 1, 2$ are 0.004 (< 0.001) and < 0.001 (< 0.001), suggesting that ZDV+ddI and ZDV+ddI+NVP improve CD4 counts for patients who have not yet developed the codon 215 drug resistance mutation.

In conclusion, the analyses suggest that switching from ZDV monotherapy to combination therapy improves the CD4 cell count marker of HIV progression for subjects who have not yet had the T215Y/F drug resistance mutation, but treatment switching has little effect after the mutation developed. The results are consistent with the result of Chapter 2.

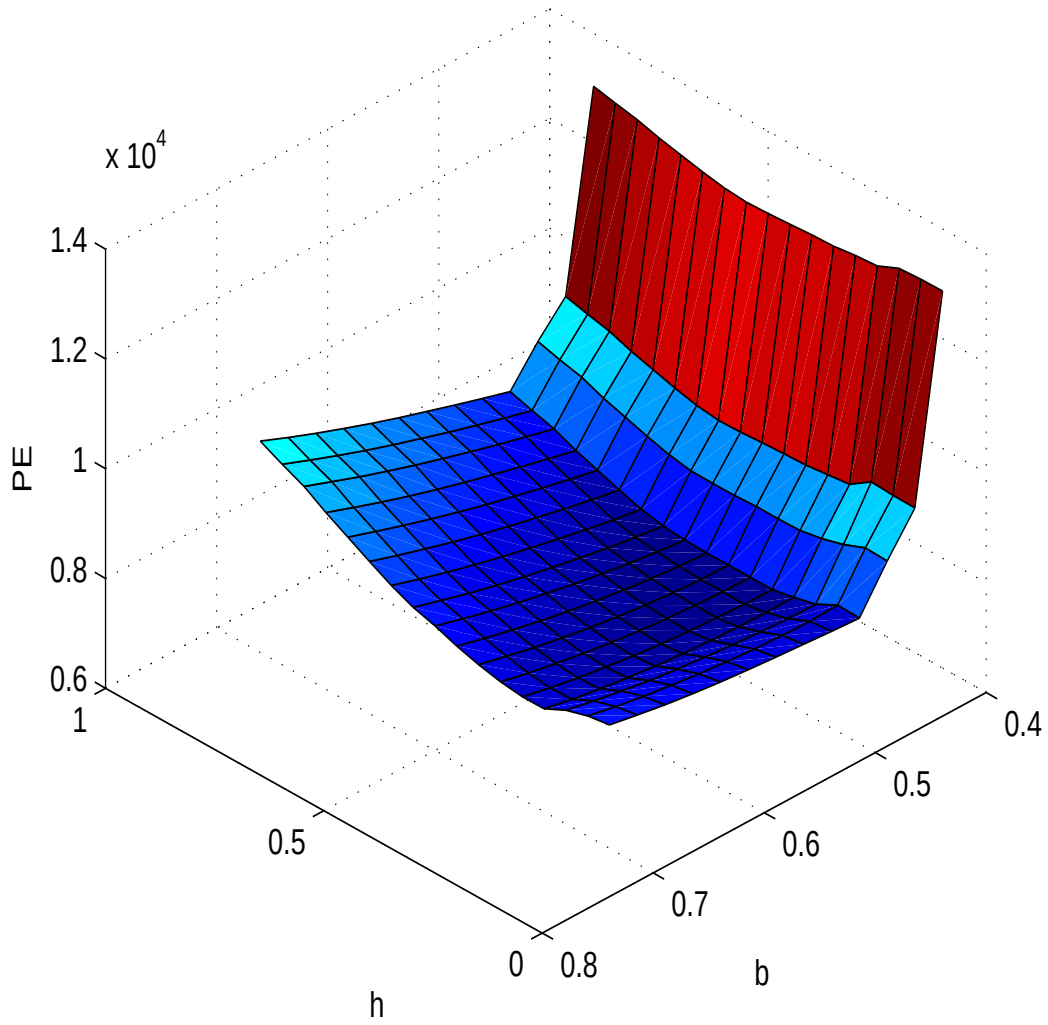


Figure 10: A preliminary study to choose suitable bandwidth for simulation with $n = 200$ and the logarithm link. The plot indicates that the optimal bandwidth are around $h = 0.45$ and $b = 0.475$.

Table 11: Summary of Bias, SEE, ESE and CP for β, and RMSEs for $\alpha(t)$ and $\gamma(u)$ under model (3.11) with identity link function.

n	h	b	Bias	SEE	ESE	CP	RMSE_α	RMSE_γ
200	0.4	0.4	0.0081	0.2640	0.2556	0.936	0.2043	0.1867
	0.5	0.5	0.0043	0.2721	0.2607	0.932	0.2103	0.1789
	0.6	0.6	-0.0161	0.2705	0.2584	0.934	0.2003	0.1738
400	0.4	0.4	-0.0068	0.1875	0.1837	0.940	0.1464	0.1309
	0.5	0.5	0.0042	0.1890	0.1842	0.936	0.1384	0.1213
	0.6	0.6	-0.0145	0.1849	0.1848	0.956	0.1378	0.1157
600	0.4	0.4	-0.0121	0.1502	0.1512	0.966	0.1162	0.1086
	0.5	0.5	0.0047	0.1537	0.1512	0.942	0.1178	0.0955
	0.6	0.6	-0.0032	0.1601	0.1520	0.942	0.1196	0.0925

Table 12: Summary of Bias, SEE, ESE and CP for β, and RMSEs for $\alpha(t)$ and $\gamma(u)$ under model (3.12) with logarithm link function.

n	h	b	Bias	SEE	ESE	CP	RMSE_α	RMSE_γ
200	0.4	0.4	0.0022	0.0633	0.0604	0.930	0.0822	0.0753
	0.5	0.5	0.0042	0.0898	0.0631	0.936	0.0826	0.0645
	0.6	0.6	-0.0018	0.1276	0.0656	0.944	0.0777	0.0633
400	0.4	0.4	0.0014	0.0428	0.0430	0.952	0.0581	0.0527
	0.5	0.5	-0.0023	0.0459	0.0444	0.940	0.0573	0.0470
	0.6	0.6	-0.0032	0.0456	0.0462	0.950	0.0544	0.0447
600	0.4	0.4	0.0004	0.0362	0.0351	0.946	0.0484	0.0410
	0.5	0.5	0.0000	0.0349	0.0364	0.952	0.0448	0.0386
	0.6	0.6	0.0006	0.0356	0.0378	0.956	0.0466	0.0347

Table 13: Summary of Bias, SEE, ESE and CP for β, and RMSEs for $\alpha(t)$ and $\gamma(u)$ under model (3.13) with logit link function.

n	h	b	Bias	SEE	ESE	CP	RMSE_α	RMSE_γ
200	0.4	0.4	0.0003	0.2308	0.1706	0.934	0.3012	0.3012
	0.5	0.5	-0.0058	0.1741	0.1723	0.932	0.2717	0.2544
	0.6	0.6	-0.0030	0.1928	0.1743	0.926	0.2509	0.2172
400	0.4	0.4	0.0000	0.1192	0.1170	0.930	0.2045	0.2101
	0.5	0.5	0.0048	0.1224	0.1186	0.924	0.1834	0.1705
	0.6	0.6	-0.0113	0.1331	0.1205	0.910	0.1804	0.1507
600	0.4	0.4	0.0065	0.0987	0.0978	0.956	0.1630	0.1698
	0.5	0.5	-0.0019	0.0986	0.0977	0.946	0.1510	0.1366
	0.6	0.6	-0.0019	0.1005	0.0997	0.940	0.1404	0.1211

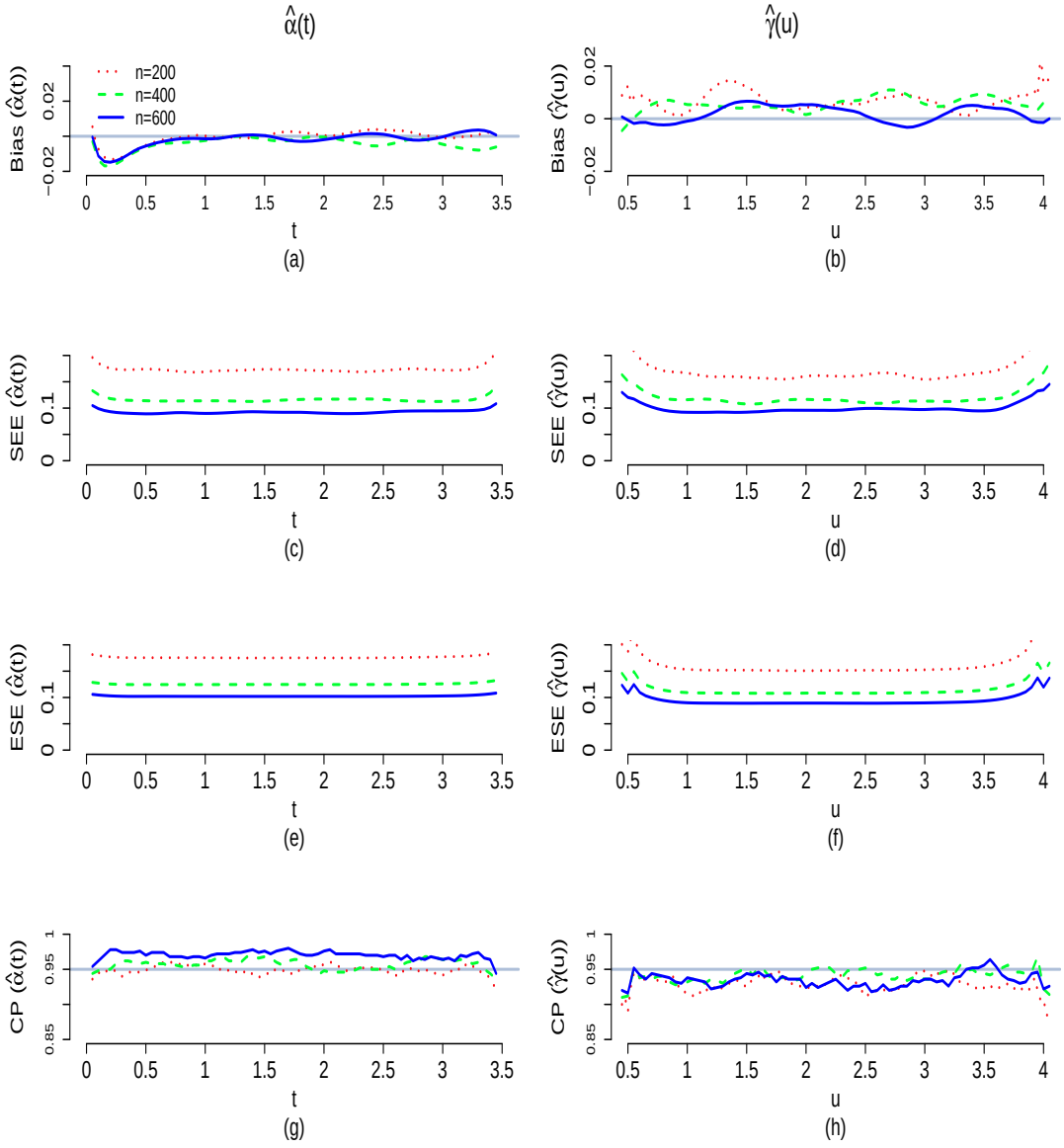


Figure 11: Plots for bias, SEE, ESE and CP for n=200, 400, 600 for identity link with h=0.4,b=0.4. Left panel is for $\alpha_0(t) = .5\sqrt{t}$. Right panel is for $\gamma(u) = -.6u$.

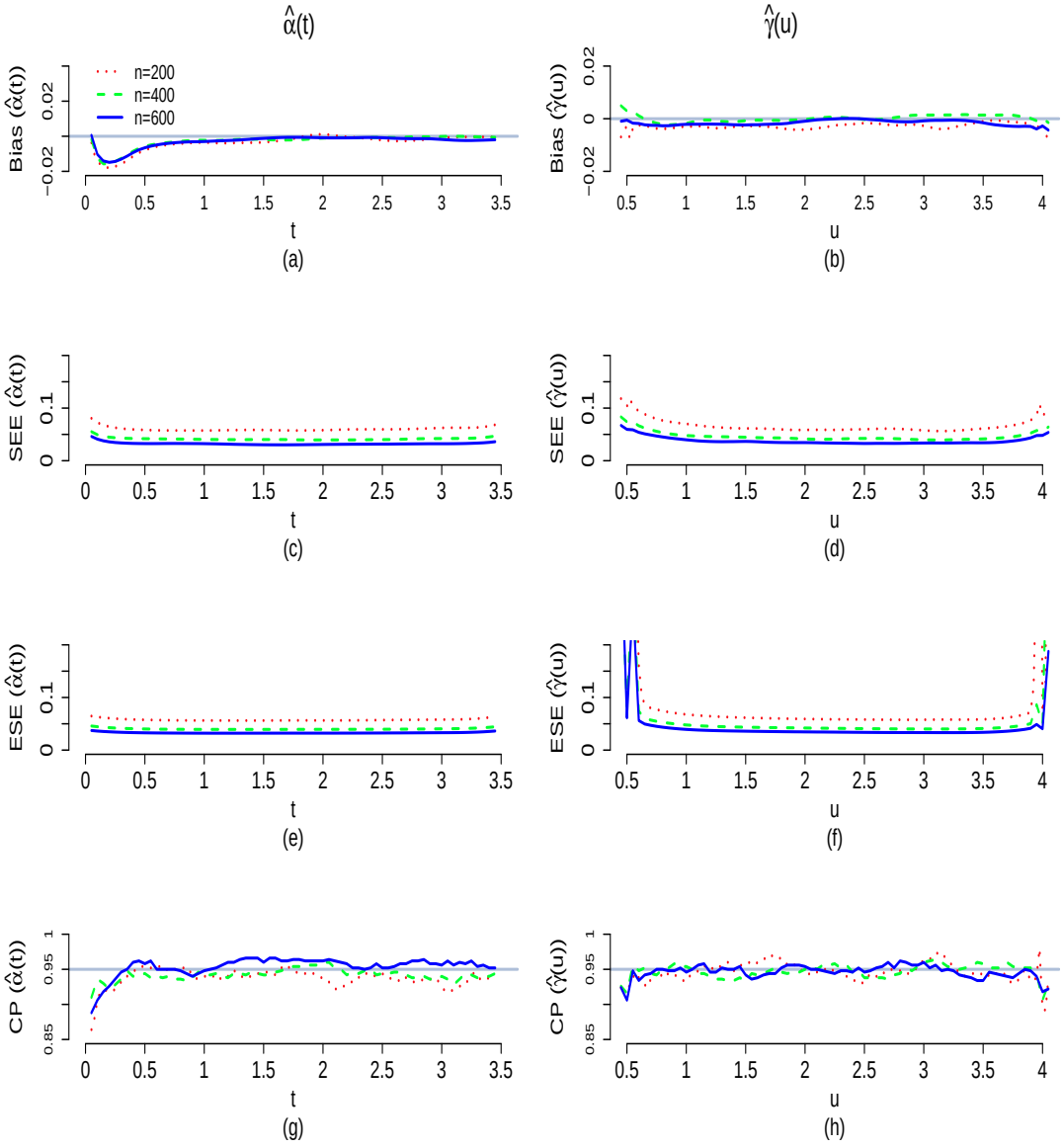


Figure 12: Plots for bias, SEE, ESE and CP for n=200, 400, 600 for logarithm link with h=0.4,b=0.4. Left panel is for $\alpha_0(t) = .5\sqrt{t}$. Right panel is for $\gamma(u) = -.6u$.

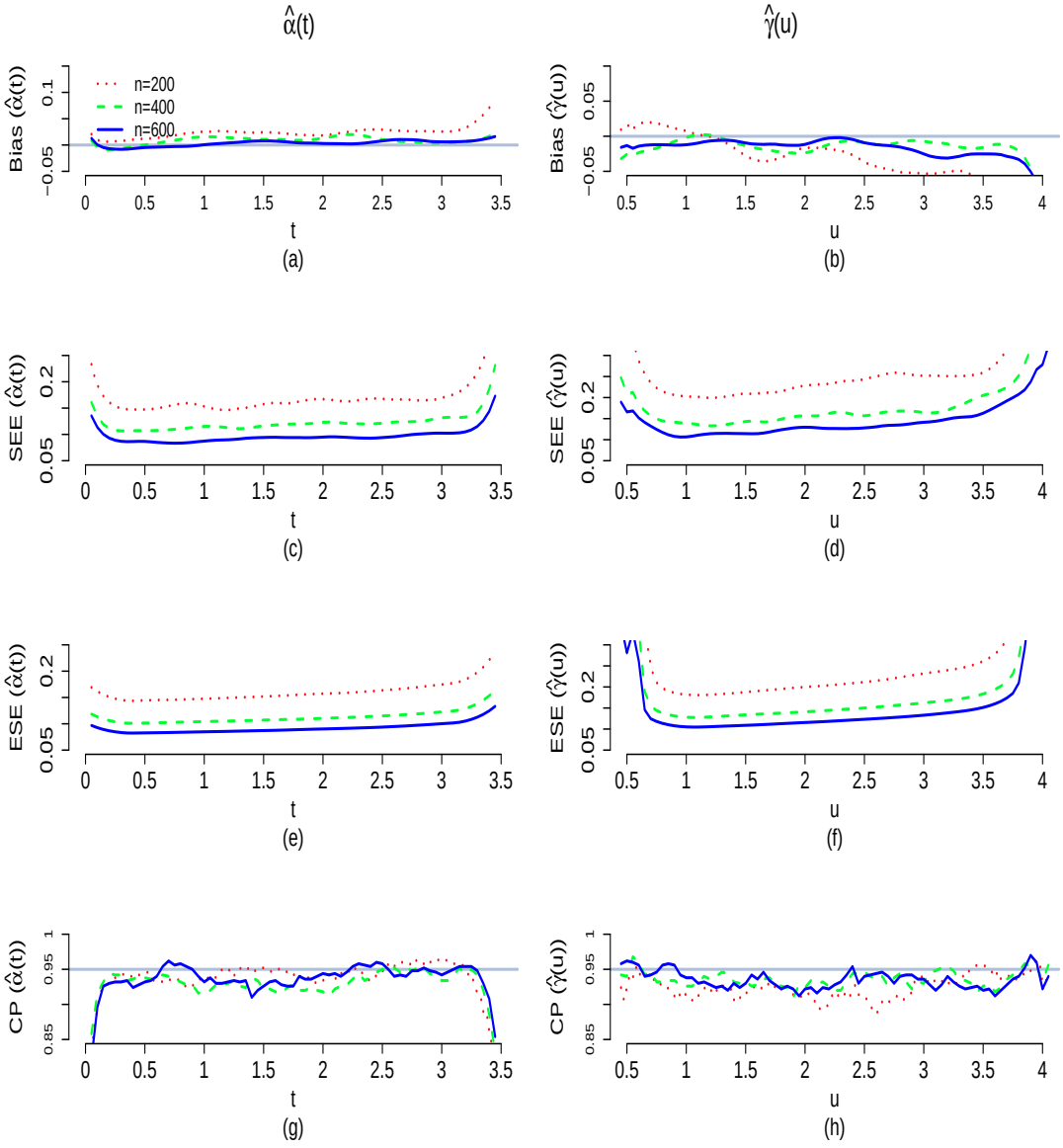


Figure 13: Plots for bias, SEE, ESE and CP for n=200, 400, 600 for logit link with h=0.4,b=0.4. Left panel is for $\alpha_0(t) = .5\sqrt{t}$. Right panel is for $\gamma(u) = -.6u$.

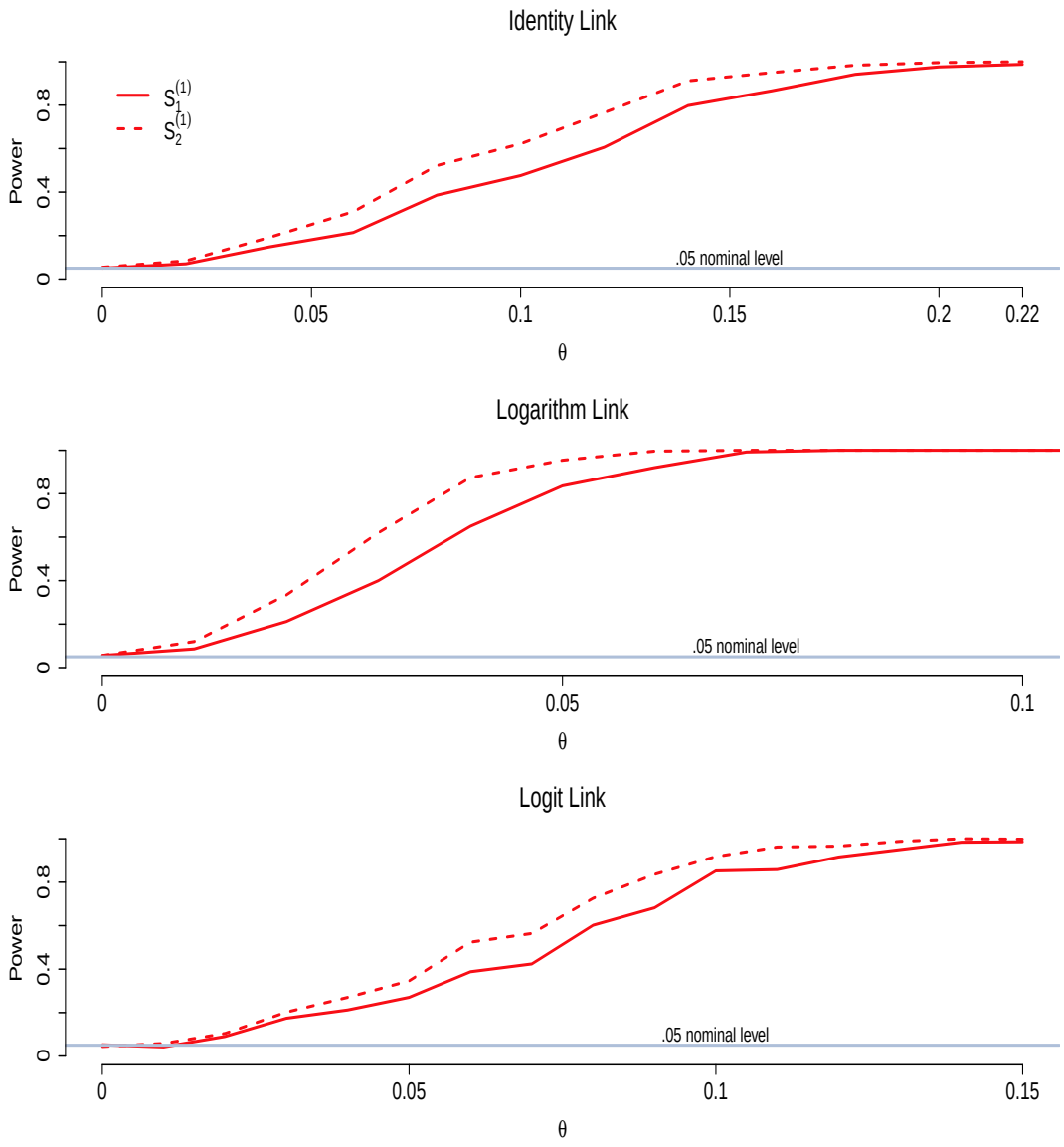


Figure 14: The power curves of the test for testing $H_0^{(1)}$: $\gamma(u) = 0$ for $u \in [u_1, u_2]$ against $H_a^{(1)}$: $\gamma(u) \neq 0$ for some u, with n=400 for identity link function, log link function and logit link function, based on 500 simulations.

Table 14: Point and 95% confidence interval estimates of β_1, β_2, β_3 and β_4 for model (3.14) based on the ACTG 244 data using $h = 0.5$, $b = 2.5$.

	Estimate	Standard deviation	95% Confidence limits		p-value
β_1	-1.0510	0.6617	-2.3478	0.2458	0.1122
β_2	0.0642	0.0343	-0.0029	0.1314	0.0607
β_3	-0.4178	0.8255	-2.0357	1.2001	0.6128
β_4	-0.7539	0.9259	-2.5686	1.0608	0.4155

Table 15: Point and 95% confidence interval estimates of β_1, β_2, β_3 and β_4 for model (3.15) based on the ACTG 244 data using $h = 0.5$, $b = 1.5$.

	Estimate	Standard deviation	95% Confidence limits		p-value
β_1	-0.4899	0.9355	-2.3236	1.3437	0.6005
β_2	0.0819	0.0358	0.0117	0.1521	0.0221
β_3	1.2564	1.0559	-0.8131	3.3259	0.2341
β_4	0.7582	1.1206	-1.4381	2.9546	0.4986

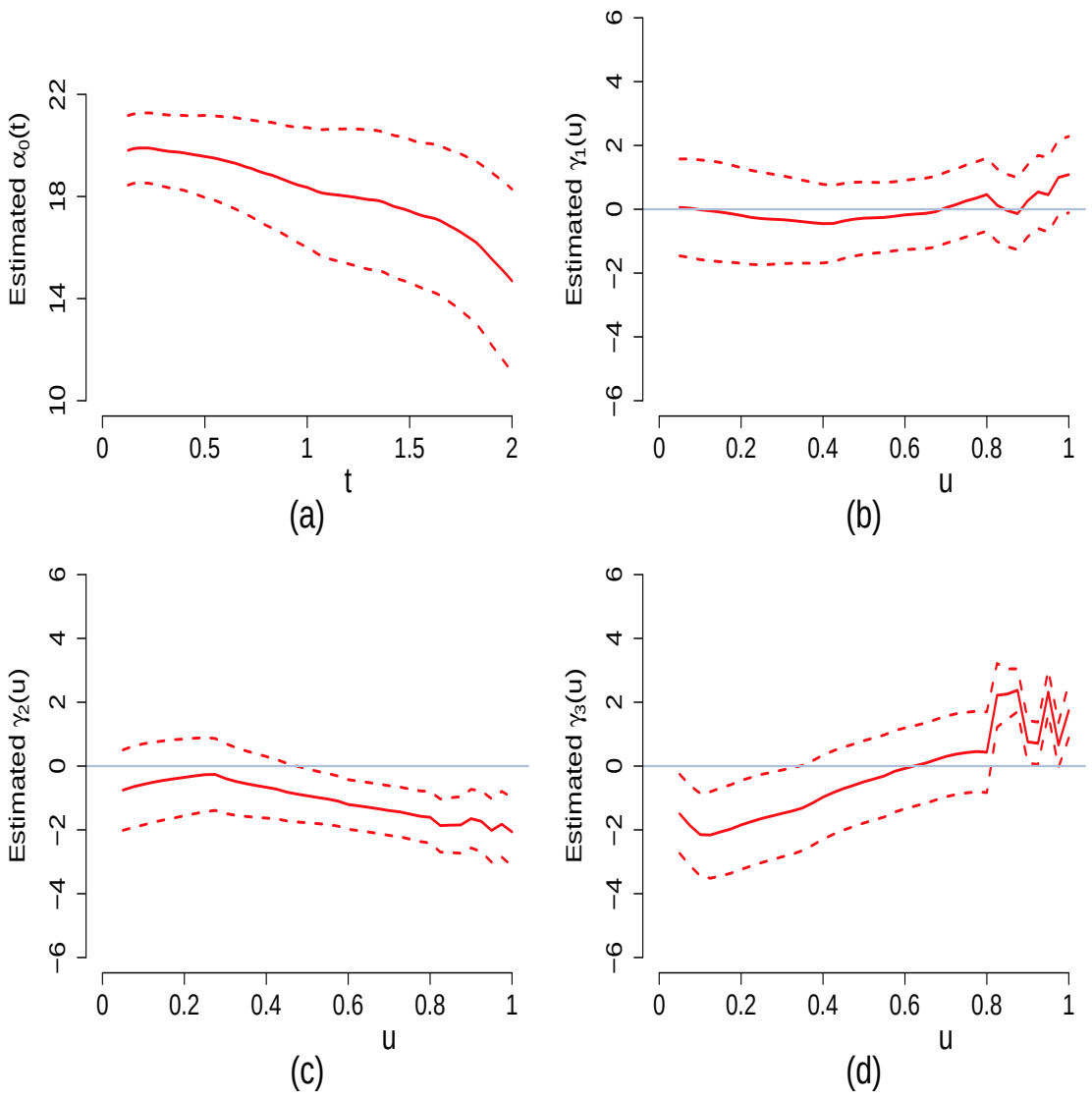


Figure 15: Plots of $\hat{\alpha}_0(t)$, $\hat{\gamma}_k(u)$, $k = 1, 2, 3$ with their 95% pointwise confidence intervals under model (3.14) based on the ACTG 244 data using $h = 0.5$ and $b = 2.5$.

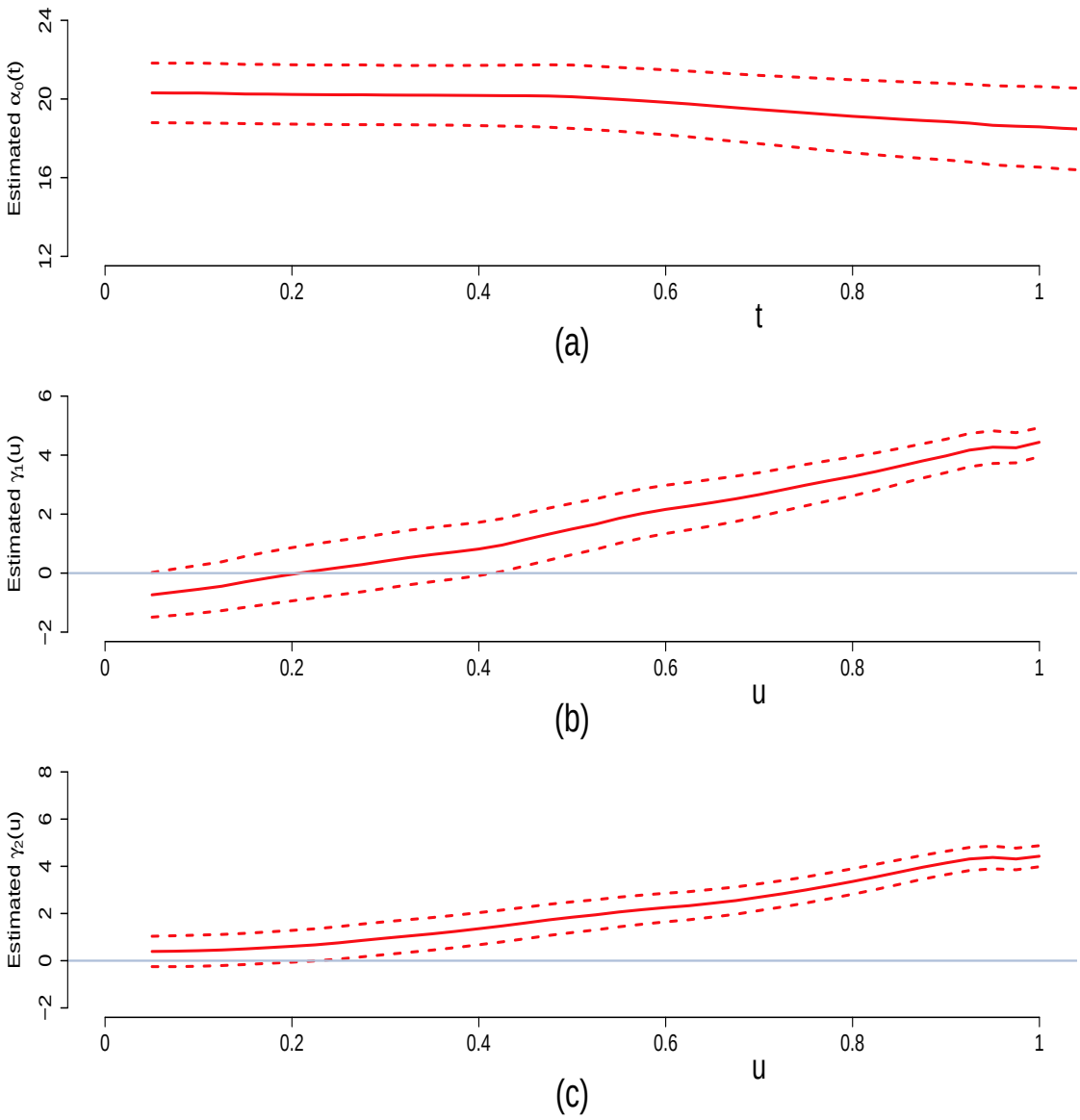


Figure 16: Plots of $\hat{\alpha}_0(t)$, $\hat{\gamma}_k(u)$, $k = 1, 2$ with their 95% pointwise confidence intervals under model (3.15) based on the ACTG 244 data using $h = 0.5$ and $b = 1.5$.

CHAPTER 4: DATA EXAMPLE: STEP STUDY WITH MITT CASES

All previous analyses of HIV vaccine efficacy trials assessed the biomarkers based on the time from the diagnosis with Ab+. While active treatments start from the time of diagnosis, it is biologically meaningful to assess whether and how vaccination modifies or accelerates the development of these biomarkers over time since the actual HIV acquisition. The time of actual HIV acquisition can be approximated well with more advanced PCR test for patients shown Ab+. Hence two time-scales are involved, one is the time from diagnosis from which time patients may start antiretroviral treatments and the longitudinal biomarkers, e.g., viral loads and CD4 counts are regularly monitored. The other one is the time from actual HIV acquisition. Simultaneous modeling of the two time-scales enable understanding the effects of treatments that started from the time of diagnosis as well as the possible time-dependent confounding between the treatments and vaccinations.

The proposed methods can be applied to solve such two-time-scale problems. A multi-center, double-blind, randomized, placebo-controlled, phase II test-of-concept STEP study (cf. Buchbinder et al. (2008); Fitzgerald et al. (2011)) was to determine whether the MRKAd5 HIV-1 gag/pol/nef vaccine, which elicits T cell immunity, is capable to result in controlling the replication of the Human immunodeficiency virus among the participants who got HIV-infected after vaccination.

This study opened in December 2004 and was conducted at 34 sites in North Amer-

ica, the Caribbean, South America, and Australia. Three thousand HIV-1 negative participants aged from 18 to 45 who were at high risk of HIV-infection were enrolled and randomly assigned to receive vaccine or placebo in ratio 1:1, stratified by sex, study site and adenovirus type 5 (Ad5) antibody titer at baseline. Some of the participants were fully adherent to vaccinations while others not.

The analysis in this section includes a subset of the 3000 participants which involves all 174 MITT cases as of September 22, 2009. MITT cases stand for modified intention-to-treat subjects who became HIV infected during the trial. The modified intention-to-treat refers to all randomized subjects, excluding the few that were found to be HIV infected at entry. It is recommended to study males only for the entire analysis to avoid the effect of sex since there are only 15 females that are $< 10\%$ of the sample. There were 159 HIV-infected males. Each participant had the records of the first positive diagnosis (the dates of their first positive Elisa confirmed by Western Blot or RNA) and the estimated time of the infection (determined by the dates of the first positive RNA (PCR) test)

After first positive diagnosis, 18 post-infection visits were scheduled per subject at weeks 0, 1, 2, 8, 12, 26, and every 26 weeks thereafter through week 338. However, the actual time and dates of visits may vary due to each individual. During jth visit, the ith subject received tests to have the measurements of HIV virus load and CD4 cell counts before the subject started the antiretroviral therapy (ART) or was censored. The time between the first positive Elisa and ART initiation or censoring is the right censoring time. In the analysis time is measured in years. The time since the first positive Elisa to the jth visit for ith subject is denoted by T_{ij} and the time

since the first evidence of HIV infection is denoted by $U_ij = T_{ij} + O_i$, where O_i is the gap between the first positive RNA (PCR) test and the first positive Elisa. Let Y_1 be the common logarithm of HIV virus load, Y_2 be the square root of CD4 counts, X_1 be the natural logarithm of Ad5, X_2 be the site indicator (1 if North America or Australia; 0 otherwise), X_3 be the pre-protocol indicator (1 if the subject was fully adherent to vaccinations; 0 otherwise) and X_4 be the treatment indicator (1 if the subject received vaccine; 0 if receiving placebo). Our first main interest is to see the how the effects of vaccine on the HIV virus load and CD4 counts change with time since actual infection.

In the data 159 males made a total of 791 pre-ART visits. Among them there are 156 missing in CD4 cell counts and 5 missing in HIV virus load. Since there are no missing in CD4 and virus load at the same time, we could use simple imputation method to impute the missing values. At each time point separately, we use a linear regression model linking $\log_{10}$(viral load) to square root of CD4 count (for those with data on both), and use the viral load value for those with missing data to fill in the missing CD4 cell count or predict missing virus load data by CD4 values. However, at three time points there are no complete data for conducting the linear regression model fitting; at two other points there are only one complete data which is unable to complete the linear model fitting; at another time point one predicted value of virus load is relatively far beyond the range of other values of virus load and may affect the analysis results. Therefore, we delete these six visits to get the complete data for the entire analysis.

Now in this complete data set there are 159 subjects with 785 visits. 97 Of all the

participants were in the vaccine group while 62 received the placebo. 122 subjects participate in the study in North America or Australia and the rest are residents in the other sites mentioned at the beginning of this chapter. The right censoring rate of T_{ij} is 69.81%. Figure 17 is further exploration of the observation times in different time scales. It is easy to figure out that there are few data after time point 2.5. Therefore we choose $\tau = 2.5$.

After preliminary exploration of the data, X_1, X_2 and X_3 show no evidence of varying coefficients. We propose the following models based on Chapter 2 for virus load

$$Y_{1i}(t) = \alpha_0(t) + \beta_1 X_{1i}(t) + \beta_2 X_{2i}(t) + \beta_3 X_{3i}(t) + (\theta_1 + \theta_2 U_i(t))X_{4i}(t) + \epsilon_i(t), \quad (4.1)$$

and the CD4 count:

$$Y_{2i}(t) = \alpha_0(t) + \beta_1 X_{1i}(t) + \beta_2 X_{2i}(t) + \beta_3 X_{3i}(t) + (\theta_1 + \theta_2 U_i(t))X_{4i}(t) + \epsilon_i(t). \quad (4.2)$$

By the empirical bandwidth formula proposed in Chapter 2, a possible reasonable choice of the bandwidth for this data set is 0.25. The estimates of time-varying baseline function $\hat{\alpha}_0(t)$ and their 95% pointwise confidence intervals are shown in Figure 18. The estimates of time-invariant parameters are shown in Table 16. It is shown that there are no significant effects of baseline Ad5 titer, study sites or the pre-protocol on the HIV viral load level, and study sites have significant effect on the CD4 counts. Also θ_2 is significantly positive for the CD4 model, which indicates the vaccine effect changes over time since actual infection and improves the CD4 counts for later time. However, Figure 19 shows that the overall vaccine effect is not

significant. Figure 20 shows the scatter plot of the residuals from fitting the Model (4.1) and (4.2).

From Figure 18, the estimates for the baseline functions are not smooth enough, since the obseavation are right skewed. We may choose a larger bandwidth or use the following transformation of the actual visit times $t = \log(T_{ij} + 0.05) + 3$ to make the observation times more evenly distributed. The estimates of $\alpha_0(t)$ and their 95% pointwise confidence intervals in the log transformed time scale are shown in Figure 21 and the time-invariant parameter estimations are in Table 17. The baseline virus load function is very flat while the baseline CD4 function is decreasing over time. None of the parameter estimators are significant.

Next we model the $\gamma(U_i(t))$ nonparametrically based on Chapter 3 for virus load

$$Y_{1i}(t) = \alpha_0(t) + \beta_1 X_{1i}(t) + \beta_2 X_{2i}(t) + \beta_3 X_{3i}(t) + \gamma(U_i(t)) X_{4i}(t) + \epsilon_i(t), \qquad (4.3)$$

and the CD4 count:

$$Y_{2i}(t) = \alpha_0(t) + \beta_1 X_{1i}(t) + \beta_2 X_{2i}(t) + \beta_3 X_{3i}(t) + \gamma(U_i(t)) X_{4i}(t) + \epsilon_i(t) \qquad (4.4)$$

We choose $\tau = 2.5$ and $h = 0.5$ and $b = 0.4$ by 3-fold cross-validation. The estimates of time-invariant parameters are shown in Table 18. β_2 is significant (p-value=0.0016) in Model (4.4). The estimates of $\alpha_0(t)$, $\gamma(u)$ and their 95% pointwise confidence intervals are shown in Figure 22. In Model (4.3), $\gamma(u)$ is increasing but not significant since p-values are 0.216 and 0.098 for $S_1^{(1)}$ and for $S_2^{(1)}$ separately. In Model (4.4), there is no clear trend for $\gamma(u)$ and it is not significant since p-values are 0.625 and 0.722 for $S_1^{(1)}$ and for $S_2^{(1)}$ separately.

Table 16: Summary statistics of the estimators of β_1, β_2, β_3, θ_1 and θ_2 for Model (4.1) and Model (4.2).

	Estimate	Standard deviation	95% Confidence limits		p-value
		Model (4.1): Virus load model			
β_1	0.0132	0.0415	-0.0681	0.0945	0.7505
β_2	-0.1171	0.1633	-0.4372	0.2030	0.4734
β_3	-0.0713	0.1752	-0.4146	0.2720	0.6840
θ_1	-0.0411	0.2888	-0.6071	0.5249	0.8868
θ_2	-0.0759	0.1368	-0.3440	0.1923	0.5792
		Model (4.2): CD4 model			
β_1	-0.2617	0.1743	-0.6033	0.0800	0.1333
β_2	3.1931	0.7250	1.7722	4.6141	$< .0001$
β_3	-0.5376	0.8613	-2.2258	1.1506	0.5325
θ_1	-1.7745	1.2981	-4.3188	0.7698	0.1716
θ_2	1.1132	0.5332	0.0682	2.1582	0.0368

Table 17: Summary statistics of the estimators of β_1, β_2, β_3, θ_1 and θ_2 for Model (4.1) and Model (4.2) in log transformed time scale.

	Estimate	Standard deviation	95% Confidence limits		p-value
		Model (4.1): Virus load model			
β_1	0.0239	0.0397	-0.0539	0.1016	0.5470
β_2	-0.0999	0.1592	-0.4120	0.2121	0.5301
β_3	-0.1268	0.1716	-0.4630	0.2095	0.4600
θ_1	0.6312	0.3848	-0.1231	1.3854	0.1010
θ_2	-0.2642	0.1473	-0.5529	0.0246	0.0729
		Model (4.2): CD4 model			
β_1	-0.1806	0.1834	-0.5400	0.1789	0.3249
β_2	3.2033	0.7254	1.7815	4.6252	$< .0001$
β_3	-0.1317	0.9393	-1.9727	1.7093	0.8885
θ_1	-1.2696	1.6745	-4.5516	2.0123	0.4483
θ_2	0.3918	0.5598	-0.7055	1.4891	0.4840

Table 18: Summary statistics of the estimators of β_1, β_2, β_3 for Model (4.3) and Model (4.4).

	Estimate	Standard deviation	95% Confidence limits		p-value
		Model (4.3): Virus load model			
β_1	0.0491	0.0592	-0.0669	0.1651	0.4071
β_2	-0.2021	0.2277	-0.6484	0.2442	0.3748
β_3	-0.0634	0.2408	-0.5353	0.4085	0.7922
		Model (4.4): CD4 model			
β_1	-0.1178	0.2934	-0.6929	0.4572	0.6879
β_2	3.9227	1.2405	1.4913	6.3540	0.0016
β_3	0.1481	1.3294	-2.4574	2.7537	0.9113

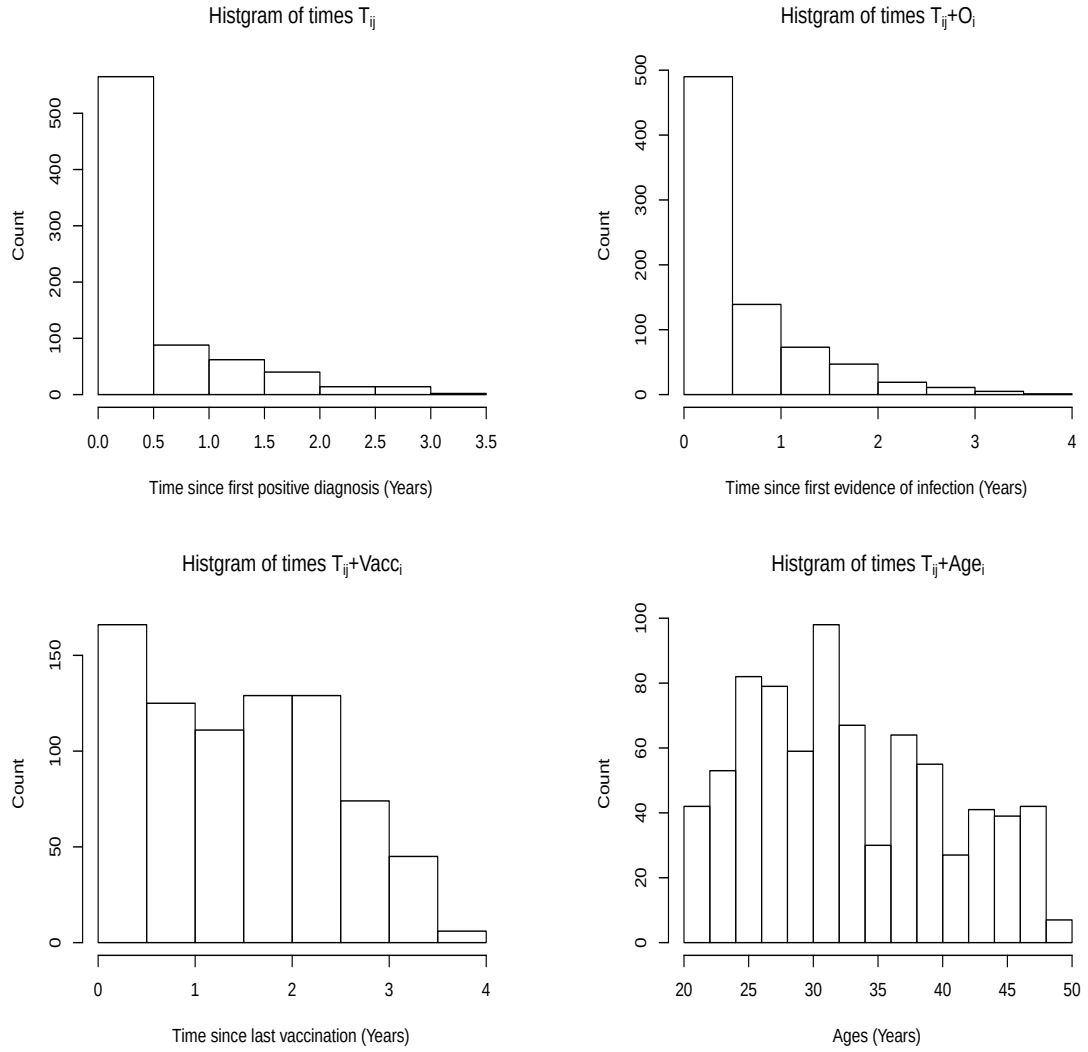


Figure 17: Histogram of several observation times in different time scales based on the data from STEP study with MITT cases.

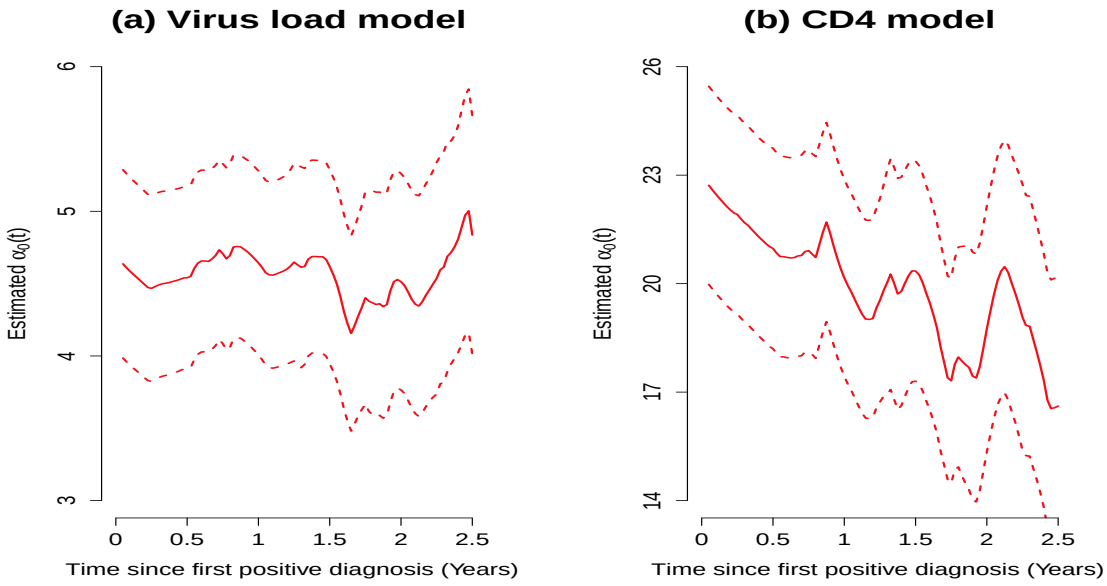


Figure 18: Estimated baseline function $\hat{\alpha}_0(t)$ and their 95% pointwise confidence intervals for Model (4.1) and Model (4.2).

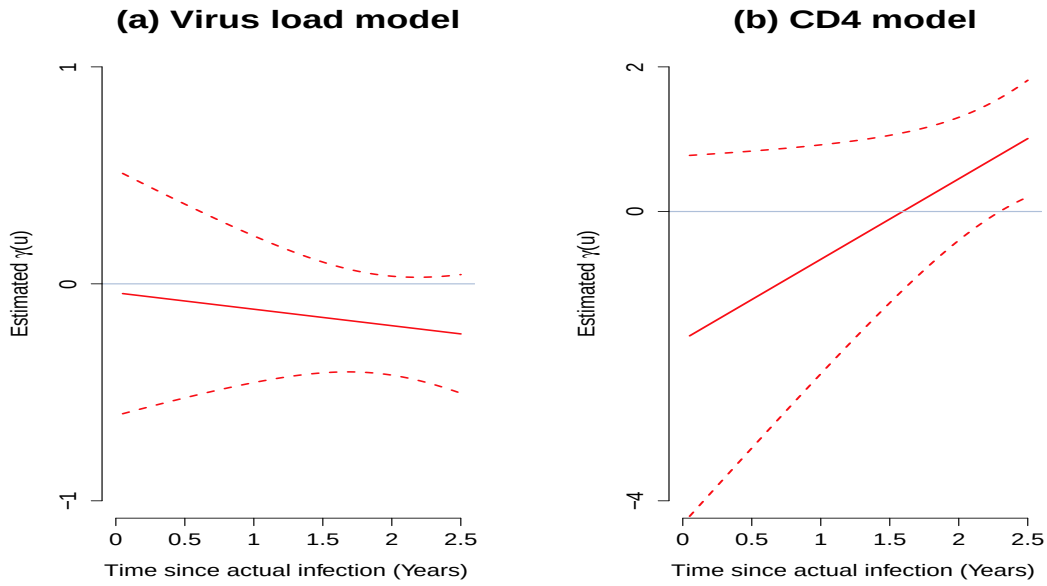


Figure 19: Estimates and the 95% confidence band of $\hat{\gamma}(u) = \hat{\theta}_1 + \hat{\theta}_2 U_i(t)$ in Model (4.2).

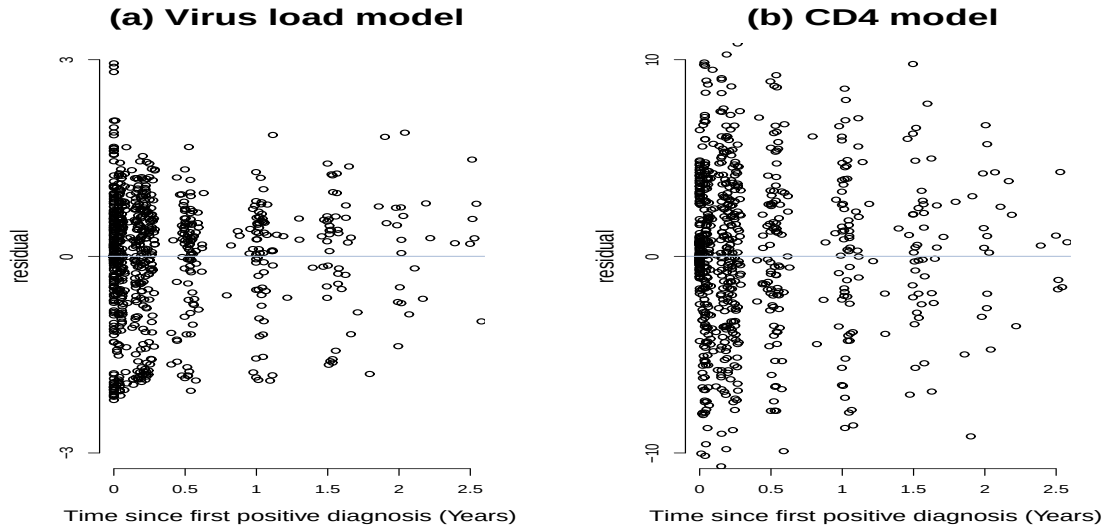


Figure 20: Scatter plots of the residuals from fitting the Model (4.1) and Model (4.2).

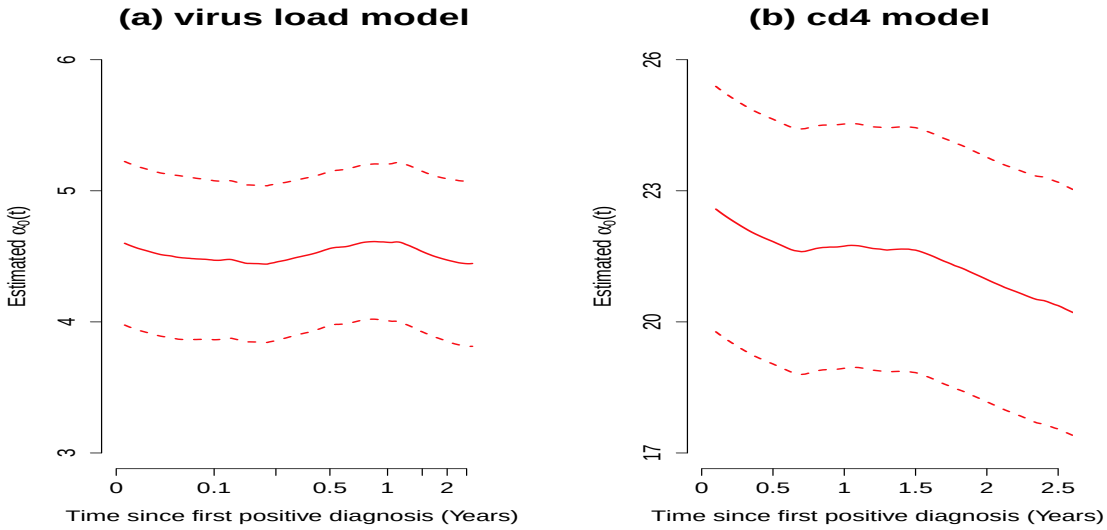
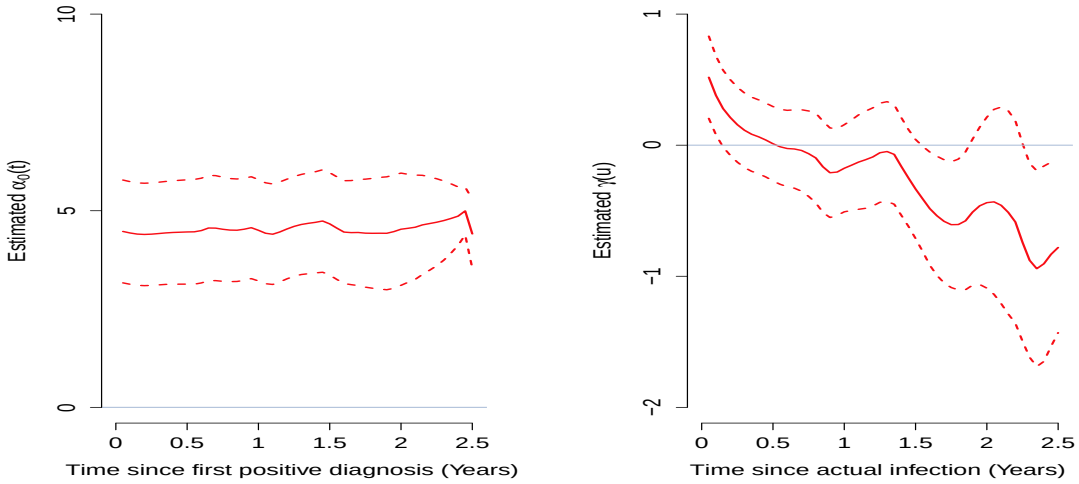


Figure 21: Estimated baseline function $\hat{\alpha}_0(t)$ and their 95% pointwise confidence intervals for Model (4.1) and Model (4.2) in log transformed time scale.

Model (4.3): Virus load model



Model (4.4): CD4 model

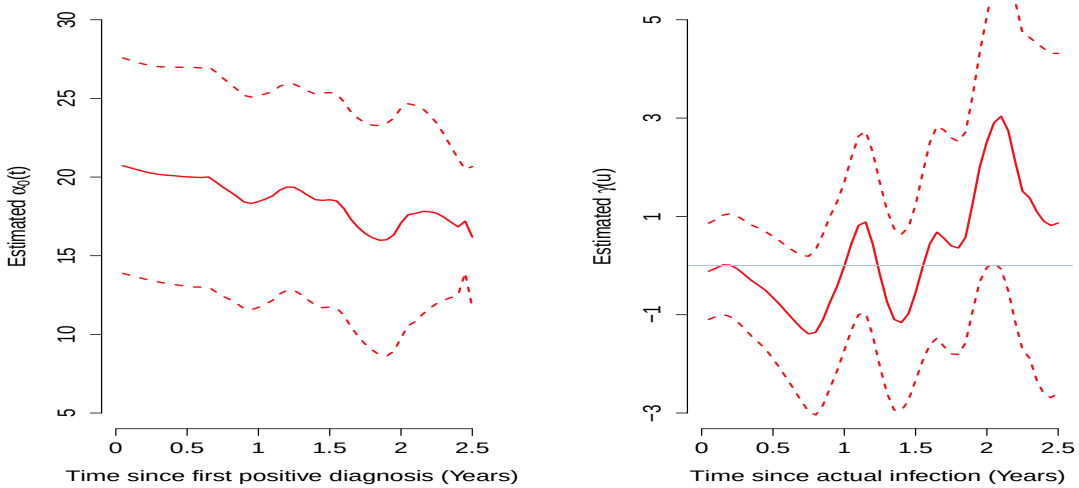


Figure 22: Estimated baseline function $\hat{\alpha}_0(t)$, $\hat{\gamma}(u)$ and their 95% pointwise confidence intervals for Model (4.3) and Model (4.4).

REFERENCES

Bickel, P. J., Klaassen, C. A., Bickel, P. J., Ritov, Y., Klaassen, J., Wellner, J. A., and Ritov, Y. (1993). *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins University Press Baltimore.

Buchbinder, S. P., Mehrotra, D. V., Duerr, A., Fitzgerald, D. W., Mogg, R., Li, D., Gilbert, P. B., Lama, J. R., Marmor, M., del Rio, C., et al. (2008). Efficacy assessment of a cell-mediated immunity hiv-1 vaccine (the step study): a double-blind, randomised, placebo-controlled, test-of-concept trial. *The Lancet*, 372(9653):1881–1893.

Cai, J. et al. (2007). Partially linear hazard regression for multivariate survival data. *Journal of the American Statistical Association*, 102(478):538–551.

Cai, J. et al. (2008). Partially linear hazard regression with varying coefficients for multivariate survival data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):141–158.

Cai, Z. and Sun, Y. (2003). Local linear estimation for time-dependent coefficients in cox's regression models. *Scandinavian Journal of Statistics*, 30(1):93–111.

Chen, Q. et al. (2013). Estimating time-varying effects for overdispersed recurrent events data with treatment switching. *Biometrika*, 100(2):339–354.

Cheng, S. and Wei, L. (2000). Inferences for a semiparametric model with panel data. *Biometrika*, 87:89–97.

Cohen, M. S. (2011). Prevention of hiv-1 infection with early antiretroviral therapy. *The New England Journal of Medicine*, 365(6):493–505.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications: Monographs on statistics and applied probability*. CRC Press.

Fan, J., Huang, T., and Li, R. (2007). Analysis of longitudinal data with semiparametric estimation of covariance function. *Journal of American Statistical Association*, 102:632–641.

Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*, 99(467).

Fan, J., Lin, H., and Zhou, Y. (2006). Local partial-likelihood estimation for lifetime data. *Annals of Statistics*, 34:290–325.

Fan, J. and Zhang, J. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society, Ser. B*, 62:303–322.

Fitzgerald, D. W. et al. (2011). An ad5- vectored hiv-1 vaccine elicitscell-mediated immunity but does not affect disease progression in hiv-1-infected male subjects: results from a randomized placebocontrolled trial (the step study). *The Journal of Infectious Diseases*, 203(6):765–772.

Gilks, C. F., Crowley, S., Ekpini, R., Gove, S., Perriens, J., Souteyrand, Y., Sutherland, D., Vitoria, M., Guerma, T., and De Cock, K. (2006). The who public-health approach to antiretroviral treatment against hiv in resource-limited settings. *The Lancet*, 368(9534):505–510.

Grabar, S., Le Moing, V., Goujard, C., Leport, C., Kazatchkine, M. D., Costagliola, D., and Weiss, L. (2000). Clinical outcome of patients with hiv-1 infection according to immunologic and virologic response after 6 months of highly active antiretroviral therapy. *Annals of internal medicine*, 133(6):401–410.

Gray, R. H. et al. (2001). Probability of hiv-1 transmission per coital act in monogamous, heterosexual, hiv-1 discordant couples in rakai, uganda. *Lancet*, 357:1149–1153.

GROUP, H. S. M. C. (2000). Human immunodeficiency virus type 1 rna level and cd4 count as prognostic markers and surrogate end points: A meta-analysis. *AIDS Research and Human Retroviruses*, 16(12):1123–1133.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 757–796.

HIV Surrogate Marker Collaborative Group (2000). Human immunodeficiency virus type 1 rna level and cd4 count as prognostic markers and surrogate endpoints: A meta-analysis. *AIDS Research and Human Retroviruses*, 16:1123–1133.

Hoover, D. R. et al. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, 85(4):809–822.

Hu, Z., W. N. and Carroll, R. J. (2004). Profile-kernel versus backfitting in the partially linear models for longitudinal/clustered data). *Biometrika*, 91(2):251–262.

Hu, X., Sun, J., and Wei, L. J. (2003). Regression parameter estimation from panel counts. *Scand J Stat.*, 30:25–43.

Huang, J. Z., Liu, N., Pourahmadi, M., and Liu, L. (2006). Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika*, 93(1):85–98.

Japour, A. J. (1995). Prevalence and clinical significance of zidovudine resistance mutations in human immunodeficiency virus isolated from patients after long-term zidovudine treatment. *J Infect Dis.*, 171(5):1172–9.

Jones, M. C., Marron, J. S., and Park, B. U. (1991). A simple root n bandwidth selector. *The Annals of Statistics*, pages 1919–1932.

Kaufmann, D., Pantaleo, G., Sudre, P., Telenti, A., Study, S. H. C., et al. (1998). Cd4-cell count in hiv-1-infected individuals remaining viraemic with highly active antiretroviral therapy (haart). *The Lancet*, 351(9104):723–724.

Larder, B. A., Kellam, P., and Kemp, S. D. (1991). Zidovudine resistance predicted by direct detection of mutations in dna from hiv-infected lymphocytes. *Aids*, 5(2):137–144.

Li, Y. (2011). Efficient semiparametric regression for longitudinal data with nonparametric covariance estimation. *Biometrika*, 98:355–370.

Lin, D. Y., Wei, L.-J., and Ying, Z. (1993). Checking the cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80(3):557–572.

Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). *Journal of the American Statistical Association*, 96:103–113.

Lin, H., Song, P. X. K., and Zhou, Q. M. (2007). Varying-coefficient marginal models and applications in longitudinal data analysis. *Sankhy's : The Indian Journal of Statistics*, 58:581–614.

Lin, X. and Carroll, R. (2000). Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American Statistical Association*, 95:520–534.

Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association*, 96:1045–1056.

Martinussen, T. and Scheike, T. H. (1999). A semiparametric additive regression model for longitudinal data. *Biometrika*, 86:691–702.

Martinussen, T. and Scheike, T. H. (2000). A nonparametric dynamic additive regression model for longitudinal data. *The Annals of Statistics*, 28:1000–1025.

Martinussen, T. and Scheike, T. H. (2001). Sampling adjusted analysis of dynamic additive regression models for longitudinal data. *Scandinavian Journal of Statistics*, 28:303–323.

Mellors, J. W., Munoz, A., Giorgi, J. V., Margolick, J. B., Tassoni, C. J., Gupta, P., Kingsley, L. A., Todd, J. A., Saah, A. J., Detels, R., et al. (1997). Plasma viral load and cd4+ lymphocytes as prognostic markers of hiv-1 infection. *Annals of internal medicine*, 126(12):946–954.

Phillips, G. D. et al. (2008). Targeting her2-positive breast cancer with trastuzumab-dm1, an antibody-cytotoxic drug conjugate. *Cancer research*, 68(22):9280–9290.

Piketty, C., Castiel, P., Belec, L., Batisse, D., Mohamed, A. S., Gilquin, J., Gonzalez-Canali, G., Jayle, D., Karmochkine, M., Weiss, L., et al. (1998). Discrepant responses to triple combination antiretroviral therapy in advanced hiv disease. *Aids*, 12(7):745–750.

Principi, N. (2001). Hiv-1 reverse transcriptase codon 215 mutation and clinical outcome in children treated with zidovudine. *AIDS Res Hum Retroviruses*, 10(6):721–6.

Quinn, T. C. et al. (2000). Viral load and heterosexual transmission of human immunodeficiency virus type 1. *New England Journal of Medicine*, 342:921–929.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society. Series B (Methodological)*, 10(6):233–243.

Scheike, T. H. (2001). A generalized additive regression model for survival times. *The Annals of Statistics*, 29:1344–1360.

Sun, J. and Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(2):293–302.

Sun, Y. (2010). Estimation of semiparametric regression model with longitudinal data. *Lifetime Data Analysis*, 16:271–298.

Sun, Y. and Gilbert, P. B. (2012). Estimation of stratified mark-specific proportional hazards models with missing marks. *Scandinavian Journal of Statistics*, 39(1):34–52.

Sun, Y., Gilbert, P. B., and McKeague, I. W. (2009a). Proportional hazards models with continuous marks. *Annals of statistics*, 37(1):394.

Sun, Y., Li, M., and Gilbert, P. B. (2013a). Mark-specific proportional hazards model with multivariate continuous marks and its application to hiv vaccine efficacy trials. *Biostatistics*, 14(1):60–74.

Sun, Y., Sun, L., and Zhou, J. (2013b). Profile local linear estimation of generalized semiparametric regression model for longitudinal data. *Lifetime Data Analysis*, 19:317–349.

Sun, Y., Sundaram, R., and Zhao, Y. (2009b). Empirical likelihood inference for the cox model with time-dependent coefficients via local partial likelihood. *Scandinavian Journal of Statistics*, 36(3):444–462.

Sun, Y. and Wu, H. (2005). Semiparametric time-varying coefficients regression model for longitudinal data. *Scandinavian Journal of Statistics*, 32:21–47.

Tian, L., Zucker, D., and Wei, L. J. (2005). On the Cox model with time-varying regression coefficients. *Journal of the American Statistical Association*, 100:172–183.

Van Der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics, 3. Cambridge University Press.

Wang, N., Carroll, R. J., and Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association*, 100:147–157.

Wu, H. and Liang, H. (2004). Backfitting random varying-coefficient models with time-dependent smoothing covariates. *Scandinavian Journal of Statistics*, 31(1):3–19.

Wu, W. B. and Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika*, 90:831–844.

Xue, L., Qu, A., and Zhou, J. (2010). Consistent model selection for marginal generalized additive model for correlated data. *Journal of the American Statistical Association*, 105:1518–1530.

Yao, F., Muller, H. G., and Wang, J. L. (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100:577–590.

Yao, F., Müller, H.-G., Wang, J.-L., et al. (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33(6):2873–2903.

Yin, G., Li, H., and Zeng, D. (2008). Partially linear additive hazards regression with varying coefficients. *Journal of the American Statistical Association*, 103(483).

Zhang, X., Park, B. U., and Wang, J. L. (2013). Time-varying additive models for longitudinal data. *Journal of the American Statistical Association*, 108:983–998.

Zhou, H. and Wang, C.-Y. (2000). Failure time regression with continuous covariates measured with error. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):657–665.

APPENDIX A: PROOFS OF THE THEOREMS IN CHAPTER 2

Condition I.

(I.1) The censoring time C_i is noninformative in the sense that $E\{dN_i^*(t)|X_i(t), U_i(t), C_i \geq t\} = E\{dN_i^*(t)|X_i(t), U_i(t)\}$ and $E\{Y_i(t)|X_i(t), U_i(t), C_i \geq t\} = E\{Y_i(t)|X_i(t), U_i(t)\}$; $dN_i^*(t)$ is independent of $Y_i(t)$ conditional on $X_i(t)$, $U_i(t)$ and $C_i \geq t$; the censoring time C_i is allowed to depend on the left continuous covariate process $X_i(\cdot)$;

(I.2) The processes $Y_i(t)$, $X_i(t)$ and $\lambda_i(t)$, $0 \leq t \leq \tau$, are bounded and their total variations are bounded by a constant;

(I.3) The kernel function $K(\cdot)$ is symmetric with compact support on $[-1, 1]$ and bounded variation; bandwidth $h \to 0$; $nh^2 \to \infty$ and nh^5 is bounded.

(I.4) $E|N_i(t_2) - N_i(t_1)|^2 \leq L(t_2 - t_1)$ for $0 \leq t_1 \leq t_2 \leq \tau$, where $L > 0$ is a constant; $E|N_i(t+h) - N_i(t-h)|^{2+v} = O(h)$, for some $v > 0$;

(I.5) The link function $g(\cdot)$ is monotone and its inverse function $g^{-1}(x)$ is twice differentiable;

(I.6) $\alpha_0(t)$, $e_{11}(t)$ and $e_{12}(t)$ are twice differentiable; $(e_{11}(t))^{-1}$ is bounded over $0 \leq t \leq \tau$; the matrices A and Σ are positive definite;

(I.7) The weight process $W(t, x) \xrightarrow{\mathcal{P}} \omega(t, x)$ uniformly in the range of (t, x); $\omega(t, x)$ is differentiable with uniformly bounded partial derivatives;

(I.8) The limit $\lim_{n\to\infty} hE\{\int_0^\tau \omega_i(s)\{Y_i(s)-\mu_i(s)\}X_{1i}(s)K_h(s-t)\,dN_i(s)\}^{\otimes 2}$ exists and is finite.

Lemmas

Let $u_\alpha(\alpha,\zeta,t) = E\{[\varphi\{\alpha_0^T(t)X_{1i}(t)+\eta^T(U_i(t),\zeta_0)X_{2i}^*(t)\}-\varphi\{\alpha^T(t)X_{1i}(t)+\eta^T(U_i(t),\zeta)X_{2i}^*(t)\}]\ X_{1i}(t)\xi_i(t)\lambda_i(t)\}$. Define $\alpha_\zeta(t)$ as the unique root such that $u_a(\alpha_\zeta,\zeta,t)=0$ for $\zeta\in\mathcal{N}_{\zeta_0}$ and $\alpha_\zeta^*(t) = (\alpha_\zeta^T(t),\mathbf{0}_{p_1}^T)^T$ where $\mathbf{0}_{p_1}$ is a $p_1\times 1$ vector of zeros. Let $e_{\zeta,11}(t) = E\big[\omega_i(t)\dot\varphi\{\alpha_\zeta^T(t)X_{1i}(t)+\eta^T(U_i(t),\zeta_0)X_{2i}^*(t)\}X_{1i}(t)^{\otimes 2}\lambda_i(t)\xi_i(t)\big]$ and $e_{\zeta,12}(t) = E\big[\omega_i(t)\dot\varphi\{\alpha_\zeta^T(t)X_{1i}(t)+\eta^T(U_i(t),\zeta_0)X_{2i}^*(t)\}X_{1i}(t)(\frac{\partial\eta(U_i(t);\zeta)}{\partial\zeta}X_{2i}^*(t))^T\lambda_i(t)\xi_i(t)\big]$. When $\zeta=\zeta_0$, we have $\alpha_\zeta(t)=\alpha_0(t)$, $e_{\zeta,11}(t)=e_{11}(t)$ and $e_{\zeta,12}(t)=e_{12}(t)$.

The following lemmas are used for proving the main theorems. The proof of the lemmas make repeated applications of the Glivenko-Cantelli Theorem (Van Der Vaart, 1998). A sufficient condition for applying the Glivenko-Cantelli Theorem can be checked by estimating the order of the bracketing number, similar to the proof of Lemma 2 of Sun et al. (2009a). This sufficient condition holds under the conditions provided in Condition I. Let $H=\mathrm{diag}\{I_{p_1},hI_{p_1}\}$.

Lemma A.1. *Under Condition I, as $n\to\infty$, $H\tilde\alpha^*(t,\zeta)\xrightarrow{\mathcal{P}}\alpha_\zeta^*(t)$,*

$$H\partial\tilde\alpha^*(t,\zeta)/\partial\zeta \xrightarrow{\mathcal{P}}(-(e_{\zeta,11}(t)^{-1}e_{\zeta,12}(t))^T,\mathbf{0}_{p_1}^T)^T, \tag{A.1}$$

and $H\partial^2\tilde\alpha^*(t,\zeta)/\partial\zeta^2$ converges in probability to a deterministic function of (t,ζ) of bounded variation, uniformly in $t\in[t_1,t_2]\subset(0,\tau)$ and $\zeta\in\mathcal{N}_{\zeta_0}$ at the rate $n^{-1/2+\nu}$ for $\nu>0$.

Proof of Lemma A.1

The first result of this lemma follows from Lemma 1 of Sun et al. (2013b) directly. We only prove the second and the third results.

By (2.6),

$$H\frac{\partial\tilde{\alpha}^*(t,\zeta)}{\partial\zeta} = -\left\{n^{-1}H^{-2}\frac{\partial U_\alpha(\alpha^*,\zeta,t)}{\partial\alpha^*}\right\}^{-1} n^{-1}H^{-1}\frac{\partial U_\alpha(\alpha^*,\zeta,t)}{\partial\zeta}\bigg|_{\alpha^*=\tilde{\alpha}^*(t,\zeta)}.$$

Note that

$$\begin{aligned}
&n^{-1}H^{-2}\frac{\partial U_\alpha(\alpha^*,\zeta,t_0)}{\partial\alpha^*}\\
&= n^{-1}\sum_{i=1}^{n}\int_0^\tau \omega_i(t)\dot{\varphi}\{\alpha^{*T}(t_0)X_{1i}^*(t,t-t_0)+\eta^T(U_i(t),\zeta)X_{2i}^*(t)\}\\
&\qquad\qquad \times H^{-2}X_i^*(t,t-t_0)^{\otimes 2}K_h(t-t_0)\,dN_i(t)\\
&= E\{\int_0^\tau \omega_i(t)\dot{\varphi}\{\alpha^{*T}(t_0)X_{1i}^*(t,t-t_0)+\eta^T(U_i(t),\zeta)X_{2i}^*(t)\}\\
&\qquad\qquad \times H^{-2}X_i^*(t,t-t_0)^{\otimes 2}K_h(t-t_0)\xi_i(t)\lambda_i(t)\,dt\} + O_p(\frac{1}{\sqrt{nh}})
\end{aligned}$$

uniformly in t by Glivenko-Cantelli Theorem.

Since $H\tilde{\alpha}^*(t,\zeta)\xrightarrow{\mathcal{P}}\alpha_\zeta^*(t)$, we have

$$\begin{aligned}
&n^{-1}H^{-2}\left.\frac{\partial U_\alpha(\alpha^*,\zeta,t_0)}{\partial\alpha^*}\right|_{\alpha^*=\tilde{\alpha}^*(t,\zeta)}\\
&= E\{\int_0^\tau \omega_i(t)\dot{\varphi}\{\alpha_\zeta^T(t_0)X_{1i}(t)+\eta^T(U_i(t),\zeta)X_{2i}^*(t)\}\\
&\qquad\qquad \times H^{-2}X_i^*(t,t-t_0)^{\otimes 2}K_h(t-t_0)\xi_i(t)\lambda_i(t)\,dt\} + O_p(\frac{1}{\sqrt{nh}})\\
&= E\{\omega_i(t_0)\dot{\varphi}\{\alpha_\zeta^T(t_0)X_{1i}(t_0)+\eta^T(U_i(t_0),\zeta)X_{2i}^*(t_0)\}\\
&\qquad\qquad \times \begin{pmatrix} 1 & 0\\ 0 & \mu_2\end{pmatrix}\otimes\{X_{1i}(t)\}^{\otimes 2}\xi_i(t_0)\lambda_i(t_0)\} + O(h^2) + O_p(\frac{1}{\sqrt{nh}})
\end{aligned}$$

uniformly in t and $\zeta \in \mathcal{N}_{\zeta_0}$. Similarly,

$$n^{-1}H^{-1}\left.\frac{\partial U_\alpha(\alpha^*,\zeta,t)}{\partial\zeta}\right|_{\alpha^*=\tilde{\alpha}^*(t,\zeta)}$$
$$\xrightarrow{\mathcal{P}}\begin{pmatrix} E[\omega_i(t_0)\dot{\varphi}\{\alpha_\zeta^T(t_0)X_{1i}(t_0)+\eta^T(U_i(t_0),\zeta)X_{2i}^*(t_0)\} \\ \qquad\qquad \times X_{1i}(t)\{\frac{\partial\eta(U_i(t);\zeta)}{\partial\zeta}X_{2i}^*(t)\}^T\xi_i(t_0)\lambda_i(t_0)] \\ 0 \end{pmatrix}$$

uniformly in t and $\zeta \in \mathcal{N}(\zeta_0)$. Therefore, (A.1) holds uniformly in t and $\zeta \in \mathcal{N}(\zeta_0)$.

By a similar argument, the third statement holds. □

Lemma A.2. *Under Condition I,*

$$\sqrt{nh}\{\tilde{\alpha}(t,\zeta_0)-\alpha_0(t)-\frac{1}{2}\mu_2h^2\ddot{\alpha}_0^T(t)\}=(e_{11}(t))^{-1}(nh)^{1/2}n^{-1}U_\alpha(\alpha_0(t),\zeta_0)+o_p(1), \tag{A.2}$$

uniformly in $t\in[t_1,t_2]\subset(0,\tau)$, *where*

$$U_\alpha(\alpha_0(t),\zeta_0)=\sum_{i=1}^n\int_0^\tau W_i(s)\epsilon_i(t)X_{1i}(s)K_h(s-t)\,dN_i(s)$$

Further, $(nh)^{1/2}n^{-1}U_\alpha(\alpha_0(t),\zeta_0)=O_p(1)$ *uniformly in* $t\in[t_1,t_2]\subset(0,\tau)$.

Proof of Lemma A.2

Applying the first order Taylor expansion to $U_\alpha(\tilde{\alpha}^*(t,\zeta_0),\zeta_0)$, we have

$$\sqrt{nh}H(\tilde{\alpha}^*(t,\zeta_0)-\alpha_0^*(t))=-\left\{n^{-1}H^{-2}\frac{\partial U_\alpha(\alpha_0^*(t))}{\partial\alpha^*}\right\}^{-1}\sqrt{\frac{h}{n}}H^{-1}U_\alpha(\tilde{\alpha}^*(t,\zeta_0),\zeta_0) \tag{A.3}$$

The first p_1 components of the above equation is

$$\sqrt{nh}(\tilde{\alpha}(t,\zeta_0)-\alpha_0(t))=-(e_{11}(t))^{-1}(h/n)^{1/2}U_\alpha(\tilde{\alpha}(t,\zeta_0),\zeta_0)\{1+o_p(1)\} \tag{A.4}$$

By the local linear approximation for $\alpha_0(t)$ around t_0,

$$\begin{aligned}
&\mu_i(t) - \varphi\{\alpha^{*T}(t_0)X_{1i}(t) + \eta^T(U_i(t),\zeta)X^*_{2i}(t)\} \\
&= \varphi\{\alpha_0^T(t)X_{1i}(t) + \eta^T(U_i(t),\zeta_0)X^*_{2i}(t)\} - \varphi\{\alpha^{*T}(t_0)X_{1i}(t) + \eta^T(U_i(t),\zeta_0)X^*_{2i}(t)\} \\
&= \dot{\mu}_i(t)\{\frac{1}{2}\ddot{\alpha}_0^T(t)X_{1i}(t)(t-t_0)^2 + O((t-t_0)^3)\}, \qquad \text{(A.5)}
\end{aligned}$$

It follows that

$$\begin{aligned}
&(h/n)^{1/2}U_\alpha(\tilde{\alpha}(t,\zeta_0),\zeta_0) \\
&= (h/n)^{1/2}\sum_{i=1}^n \int_0^\tau \omega_i(t)\left[Y_i(t) - \mu_i(t) + \mu_i(t) - \varphi\{\alpha^{*T}(t_0)X_{1i}(t) + \eta^T(U_i(t),\zeta_0)X^*_{2i}(t)\}\right] \\
&\qquad\qquad \times X_i(t)K_h(t-t_0)\,dN_i(t) \\
&= (h/n)^{1/2}\sum_{i=1}^n \int_0^\tau \omega_i(t)\epsilon_i(t)X_{1i}(t)K_h(t-t_0)\,dN_i(t) + \frac{1}{2}\mu_2\sqrt{nh}h^2\ddot{\alpha}_0^T(t)e_{11}(t)
\end{aligned}$$

Hence

$$\begin{aligned}
&\sqrt{nh}(\tilde{\alpha}(t,\zeta_0) - \alpha_0(t) - \frac{1}{2}\mu_2 h^2\ddot{\alpha}_0^T(t)) \\
&= -(e_{11}(t))^{-1}\sqrt{\frac{h}{n}}\sum_{i=1}^n \int_0^\tau W_i(s)\epsilon_i(s)X_{1i}(s)K_h(s-t)\,d\,N_i(s) + o_p((nh)^{-1} + \sqrt{nh^5})
\end{aligned}$$

Follow Appendix A of Tian et al. (2005), the right hand side of above equation is $O_p(1)$ uniformly in $t \in [t_1, t_2]$. □

Proof of Theorems

Proof of Theorem 2.1

We first consider the proof for the consistency of $\hat{\zeta}$. By Glivenko-Cantelli theorem

and Lemma A.1, we have

$$n^{-1}U_\zeta(\zeta)\xrightarrow{\mathcal{P}}E\Big\{\int_{t_1}^{t_2}\omega_i(t)\left[Y_i(t)-\varphi\{\alpha_\zeta^T(t)X_{1i}(t)+\eta^T(U_i(t),\zeta)X_{2i}^*(t)\}\right] \times\left[-(e_{\zeta,12}(t))^T(e_{\zeta,11}(t))^{-1}X_{1i}(t)+\frac{\partial\eta(U_i(t),\zeta)}{\partial\zeta}X_{2i}^*(t)\right]\,dN_i(t)\Big\}$$

uniformly for $\zeta\in\mathcal{N}_{\zeta_0}$. The right side of the above equation equals to

$$E\Big\{\int_{t_1}^{t_2}\omega_i(t)\left[\varphi\{\alpha_0^TX_{1i}(t)+\eta^T(U_i(t),\zeta_0)X_{2i}^*(t)\}-\varphi\{\alpha_\zeta^T(t)X_{1i}(t)+\eta^T(U_i(t),\zeta)X_{2i}^*(t)\}\right] \times\left[-(e_{\zeta,12}(t))^T(e_{\zeta,11}(t))^{-1}X_{1i}(t)+\frac{\partial\eta(U_i(t),\zeta)}{\partial\zeta}X_{2i}^*(t)\right]\xi_i(t)\lambda_i(t)\,dt\Big\}.$$

defined as $u(\zeta)$ by double expectation. Taking partial derivative of $U_\zeta(\zeta)$ with respect to ζ and applying Lemma 1, we have

$$\begin{aligned}
&n^{-1}\frac{\partial U_\zeta(\zeta)}{\partial\zeta}\\
&=-n^{-1}\sum_{i=1}^n\int_{t_1}^{t_2}\omega_i(t)\dot\varphi\{\tilde\alpha^T(t,\zeta)X_{1i}(t)+\eta^T(U_i(t),\zeta)X_{2i}^*(t)\}\\
&\qquad\times\left\{\frac{\partial\tilde\alpha(t,\zeta)}{\partial\zeta}X_{1i}(t)+\frac{\partial\eta(U_i(t),\zeta)}{\partial\zeta}X_{2i}^*(t)\right\}^{\otimes2}dN_i(t)\\
&\quad+n^{-1}\sum_{i=1}^n\int_{t_1}^{t_2}\omega_i(t)\left[Y_i(t)-\varphi\{\tilde\alpha^T(t,\zeta)X_{1i}(t)+\eta^T(U_i(t),\zeta)X_{2i}^*(t)\}\right]\\
&\qquad\times\left\{\frac{\partial^2\tilde\alpha(t,\zeta)}{\partial\zeta^2}X_{1i}(t)+\frac{\partial^2\eta(U_i(t),\zeta)}{\partial\zeta^2}X_{2i}^*(t)\right\}dN_i(t).
\end{aligned}\tag{A.6}$$

When $\zeta=\zeta_0$, the latter term goes to zero as n goes to infinity by Lemma A.1 and

the Glivenko-Cantelli theorem. It follows that

$$-n^{-1}\left.\frac{\partial U_\zeta(\zeta)}{\partial\zeta}\right|_{\zeta=\zeta_0} \xrightarrow{\mathcal{P}} E\{\int_{t_1}^{t_2}\omega_i(t)\dot{\varphi}\{\alpha_0^T(t)X_{1i}(t)+\eta^T(U_i(t),\zeta_0)X_{2i}^*(t)\}$$
$$\left\{-(e_{\zeta,11}(t))^{-1}e_{\zeta,12}(t)X_{1i}(t)+\frac{\partial\eta(U_i(t),\zeta)}{\partial\zeta}X_{2i}^*(t)\right\}^{\otimes 2}dN_i(t)\}$$
$$= A \tag{A.7}$$

which is positive definite, uniformly in a neighborhood of ζ_0. Since $u(\zeta_0)=0$ and A is positive, ζ_0 is the unique root of $u(\zeta)$. By Theorem 5.9 of Van Der Vaart (1998), we have the consistency of $\hat{\zeta}$.

Now we show the asymptotic normality of $n^{-1/2}U_\zeta(\zeta_0)$.

$$\begin{aligned}
&n^{-1/2}U_\zeta(\zeta_0)\\
&= n^{-1/2}\sum_{i=1}^n\int_{t_1}^{t_2}\omega_i(t)\left[Y_i(t)-\varphi\{\tilde{\alpha}^T(t,\zeta_0)X_{1i}(t)+\eta^T(U_i(t),\zeta_0)X_{2i}^*(t)\}\right]\\
&\qquad\times\left\{\frac{\partial\tilde{\alpha}(t,\zeta_0)}{\partial\zeta}X_{1i}(t)+\frac{\partial\eta(U_i(t),\zeta_0)}{\partial\zeta}X_{2i}^*(t)\right\}dN_i(t),\\
&= n^{-1/2}\sum_{i=1}^n\int_{t_1}^{t_2}\omega_i(t)\epsilon_i(t)\left\{\frac{\partial\tilde{\alpha}(t,\zeta_0)}{\partial\zeta}X_{1i}(t)+\frac{\partial\eta(U_i(t),\zeta_0)}{\partial\zeta}X_{2i}^*(t)\right\}dN_i(t), \qquad \text{(A.8)}\\
&\quad+n^{-1/2}\sum_{i=1}^n\int_{t_1}^{t_2}\omega_i(t)[\varphi\{\alpha_0^T(t)X_{1i}(t)+\eta^T(U_i(t),\zeta_0)X_{2i}^*(t)\}\\
&\qquad-\varphi\{\tilde{\alpha}^T(t,\zeta_0)X_{1i}(t)+\eta^T(U_i(t),\zeta_0)X_{2i}^*(t)\}]\\
&\qquad\times\left\{\frac{\partial\tilde{\alpha}(t,\zeta_0)}{\partial\zeta}X_{1i}(t)+\frac{\partial\eta(U_i(t),\zeta_0)}{\partial\zeta}X_{2i}^*(t)\right\}dN_i(t), \qquad \text{(A.9)}
\end{aligned}$$

The (A.9) is negligible because by Taylor expansion,

$$\varphi\{\alpha_0^T(t)X_{1i}(t)+\eta^T(U_i(t),\zeta_0)X_{2i}^*(t)\}-\varphi\{\tilde{\alpha}^T(t,\zeta_0)X_{1i}(t)+\eta^T(U_i(t),\zeta_0)X_{2i}^*(t)\}$$
$$=\dot{\mu}_i(t)\{(\tilde{\alpha}(t,\zeta_0))^T-(\alpha_0(t))^T\}X_{1i}(t)+O_p(\|\tilde{\alpha}(t,\zeta_0)-\alpha_0(t)\|^2).$$

$$
\begin{aligned}
(A.9) &= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}\omega_i(t)\dot{\mu}_i(t)\{(\tilde{\alpha}(t,\zeta_0))^T-(\alpha_0(t))^T\}X_{1i}(t)\\
&\quad\times\left\{\frac{\partial\tilde{\alpha}(t,\zeta_0)}{\partial\zeta}X_{1i}(t)+\frac{\partial\eta(U_i(t),\zeta_0)}{\partial\zeta}X_{2i}^{*}(t)\right\}dN_i(t),\\
&= o_p(1)
\end{aligned}
$$

by Lemma 1 in Lin and Ying (2001).

Hence,

$$
\begin{aligned}
&n^{-1/2}U_{\zeta}(\zeta_0)\\
&= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}\omega_i(t)\epsilon_i(t)\left\{\frac{\partial\tilde{\alpha}(t,\zeta_0)}{\partial\zeta}X_{1i}(t)+\frac{\partial\eta(U_i(t),\zeta_0)}{\partial\zeta}X_{2i}^{*}(t)\right\}dN_i(t)+o_p(1)\\
&= n^{-1/2}\sum_{i=1}^{n}\int_{t_1}^{t_2}\omega_i(t)\epsilon_i(t)\left\{(e_{11}(t))^{-1}e_{12}(t)X_{1i}(t)+\frac{\partial\eta(U_i(t),\zeta_0)}{\partial\zeta}X_{2i}^{*}(t)\right\}dN_i(t)+o_p(1),
\end{aligned}
\tag{A.10}
$$

which converges in distribution to $N(0,\Sigma)$ by central limit theorem, where Σ_{ζ} is defined in (2.9).

It follows from (A.7) and (A.10) that $n^{1/2}(\hat{\zeta}-\zeta_0)\xrightarrow{\mathcal{D}}N(0,A^{-1}\Sigma_{\zeta}A^{-1})$.

Proof of Theorem 2.2

Since $\hat{\alpha}(t_0)=\tilde{\alpha}(t_0,\hat{\zeta})$, we have $\hat{\alpha}(t_0)\xrightarrow{\mathcal{P}}\alpha_0(t_0)$ uniformly in $t\in[t_1,t_2]$ by applying continuous mapping theorem and the uniform consistency results in Lemma A.1 and Theorem 2.1. Now we prove the asymptotic normality.

By Taylor expansion we have

$$
\sqrt{nh}(\tilde{\alpha}(t_0,\hat{\zeta})-\tilde{\alpha}(t,\zeta_0))=-(nh)^{1/2}\frac{\partial\tilde{\alpha}(t_0,\tilde{\zeta})}{\partial\zeta}(\hat{\zeta}-\zeta_0)
$$

where $\tilde{\zeta}$ is on the line segment between ζ_0 and $\hat{\zeta}$, which is $O_p(h^{1/2})$, by (A.1) and

Theorem 2.1. Thus

$$\begin{aligned}
&\sqrt{nh}\{\hat{\alpha}(t) - \alpha_0(t) - \frac{1}{2}\mu_2 h^2 \ddot{\alpha}_0^T(t)\} \\
&= \sqrt{nh}\{\tilde{\alpha}(t,\zeta_0) - \alpha_0(t) - \frac{1}{2}\mu_2 h^2 \ddot{\alpha}_0^T(t)\} + \sqrt{nh}(\tilde{\alpha}(t_0,\hat{\zeta}) - \tilde{\alpha}(t,\zeta_0)) \\
&= (e_{11}(t))^{-1}(h/n)^{1/2} \sum_{i=1}^{n} \int_0^\tau \omega_i(s)\{Y_i(s) - \mu_i(s)\} X_i(s) K_h(s-t)\, d\, N_i(s) + O_p(h^{1/2}), \\
&= (e_{11}(t))^{-1} n^{-1/2} \sum_{i=1}^{n} \psi_i(t) + O_p(h^{1/2}),
\end{aligned}$$

for $t \in [t_1, t_2]$ by (A.2).

Note that $E(\psi_i(t)) = 0$. It follows that $n^{-1/2}\sum_{i=1}^n \psi_i(t) \xrightarrow{\mathcal{D}} N(0, \Sigma_e)$ by applying the Lindeberg-Feller central limit theorem. Consequently,

$$\sqrt{nh}(\hat{\alpha}(t) - \alpha_0(t) - \frac{1}{2}\mu_2 h^2 \ddot{\alpha}_0^T(t)) \xrightarrow{\mathcal{D}} N\left(0, (e_{11}(t))^{-1}\Sigma_e(t)(e_{11}(t))^{-1}\right).$$

APPENDIX B: PROOFS OF THE THEOREMS IN CHAPTER 3

Condition II.

(II.1) The censoring time C_i is noninformative in the sense that $E\{dN_i^*(t)|X_i(t), U_i(t), C_i \geq t\} = E\{dN_i^*(t)|X_i(t), U_i(t)\}$ and $E\{Y_i(t)|X_i(t), U_i(t), C_i \geq t\} = E\{Y_i(t)|X_i(t), U_i(t)\}$; $dN_i^*(t)$ is independent of $Y_i(t)$ conditional on $X_i(t)$, $U_i(t)$ and $C_i \geq t$; the censoring time C_i is allowed to depend on the left continuous covariate process $X_i(\cdot)$;

(II.2) The processes $Y_i(t)$, $X_i(t)$ and $\lambda_i(t)$, $0 \leq t \leq \tau$, are bounded and their total variations are bounded by a constant; $E|N_i(t_2) - N_i(t_1)|^2 \leq L(t_2 - t_1)$ for $0 \leq t_1 \leq t_2 \leq \tau$, where $L > 0$ is a constant; $E|N_i(t+h) - N_i(t-h)|^{2+v} = O(h)$, for some $v > 0$;

(II.3) The kernel function $K(\cdot)$ is symmetric with compact support on $[-1, 1]$ and Lipschitz contimuous; Bandwidths $h \asymp b$; $h \to 0$; $nh^2 \to \infty$ and nh^5 is bounded;

(II.4) The function $\varphi(\cdot)$ is monotone twice differentiable;

(II.5) $\alpha_0(t)$, $\gamma_0(u)$, $e_{11}(t)$ and $e_{12}(t)$ are twice differentiable; $(e_{11}(t))^{-1}$ is bounded over $0 \leq t \leq \tau$; the matrices A and Σ are positive definite;

(II.6) The limit $\lim_{n\to\infty} hE\left[\int_0^\tau [Y_i(t) - \mu_i(t)]\, \mathscr{I}_1 e_{11}(t_0, U_i(t))^{-1} X_i(t) K_h(t - t_0)\, dN_i(t)\right]^{\otimes 2}$, and

$\lim_{n\to\infty} bE\left[\int_0^\tau \{Y_i(t) - \hat{\mu}_i(t)\}\, \mathscr{I}_3 e_{11}(t, u_0)^{-1} X_i(t) K_b(U_i(t) - u_0)\, dN_i(t)\right]^{\otimes 2}$ exist and are finite.

Lemmas

Let random function $\psi : \mathbb{R}^4 \to \mathbb{R}$ satisfy: $\psi(t, y, x, u)$ are continuous on $\{(t, x, u)\}$, uniformly in $y \in \mathbb{R}$; $E|\psi|^s < \infty$ for $s > 2$. Let $\psi_i(t) = \psi(t, Y_i(t), X_i(t), U_i(t))$. The kernel-weighted averages for two-dimensional smoothers are defined as

$$\Psi_n(t_0, u_0) = n^{-1} \sum_{i=1}^{n} \int_{t_1}^{t_2} \psi_i(t) K_h(t - t_0) K_b(U_i(t) - u_0)\, dN_i(t).$$

Lemma B.1. *Under some conditions II, and assume $h \asymp b$; $h \to 0$; $\sqrt{nh}b^2 \to \infty$, we have*

$$\begin{aligned} &\sup_{t_0 \in [t_1, t_2], u_0 \in [u_1, u_2]} |\Psi_n(t_0, u_0) - E\{\psi_i(t_0)\xi_i(t_0)\lambda_i(t_0)|U(t_0) = u_0\} f_U(t_0, u_0)| \\ &= O_p(\log n/\sqrt{nhb} + h^2 + b^2) \end{aligned}$$

Proof of Lemma B.1

Let $M_i(t) = N_i(t) - \int_0^t \xi_i(s)\lambda_i(s)ds$ be a mean zero stochastic process, we have

$$\Psi_n(t_0, u_0) = n^{-1} \sum_{i=1}^{n} \int_{t_1}^{t_2} \psi_i(t) K_h(t - t_0) K_b(U_i(t) - u_0) \{dM_i(t) + \xi_i(t)\lambda_i(t)\, dt\}$$

Following the argument in Lemma 1 of Zhang et al. (2013), we have that

$$\sup_{t_0 \in [t_1, t_2], u_0 \in [u_1, u_2]} \left| n^{-1} \sum_{i=1}^{n} \int_{t_1}^{t_2} \psi_i(t) K_h(t - t_0) K_b(U_i(t) - u_0)\, dM_i(t) \right| = O_p(\log n/\sqrt{nhb}).$$

Then

$$\Psi_n(t_0, u_0) = n^{-1} \sum_{i=1}^{n} \int_{t_1}^{t_2} \psi_i(t) K_h(t - t_0) K_b(U_i(t) - u_0)\xi_i(t)\lambda_i(t)\, dt + O_p(\log n/\sqrt{nhb}).$$

uniformly in $t_0 \in [t_1, t_2]$, $u_0 \in [u_1, u_2]$. Note that the first term of above is equal to

$$\int_{t_1}^{t_2} K_h(t - t_0) n^{-1} \sum_{i=1}^{n} \{\psi_i(t) K_b(U_i(t) - u_0) \xi_i(t) \lambda_i(t)\}\, dt. \tag{B.1}$$

By applying Lemma A.1 in Yin et al. (2008) , (B.1) becomes

$$\int_{t_1}^{t_2} K_h(t - t_0)[E\{\psi_i(t)\xi_i(t)\lambda_i(t)|U(t) = u_0\} f_{U(t)}(u_0) + O_p(\log n/\sqrt{n}b + b^2)]\, dt,$$

and by the Taylor series expansion, it follows that

$$\Psi_n(t_0, u_0) = E\{\psi_i(t_0)\xi_i(t_0)\lambda_i(t_0)|U(t_0) = u_0\} f_U(t_0, u_0) + O_p(\log n/\sqrt{nhb} + h^2 + b^2).$$

□

Lemma B.2. *Let Θ and $\mathcal{U}$ be compact sets in R^p and R^q, and let $\Phi_n(\theta, u)$ be random functions and let $\Phi(\theta, u)$ be a fixed function of $(\theta, u) \in \Theta \times \mathcal{U}$. Let $\delta(u)$ be a fixed function of $u \in \mathcal{U}$ taking values in Θ. Assume that $\sup_{\theta,u} \parallel \Phi_n(\theta, u) - \Phi(\theta, u) \parallel \xrightarrow{\mathcal{P}} 0$ and that for every $\varepsilon > 0$, we have $\inf_{\|\theta - \delta_0(u)\| > \varepsilon} \parallel \Phi(\theta, u) \parallel > 0 = \parallel \Phi(\delta_0(u), u) \parallel$ for $u \in \mathcal{U}$. Then for any sequence of estimators $\hat{\delta}(u)$, with $\Phi_n(\hat{\delta}(u), u) = o_p(1)$ uniformly in $u \in \mathcal{U}$, we have $\hat{\delta}(u) \xrightarrow{\mathcal{P}} \delta_0(u)$ uniformly in $u \in \mathcal{U}$.*

Proof of Lemma B.2

This follows from the Lemma 1 of Sun et al. (2009a), on applying it to the functions $\Phi_n(u, \theta) = - \parallel Q_n(u, \theta) \parallel$ and $\Phi(u, \theta) = - \parallel Q(u, \theta) \parallel$. □

Let H is a $(2p_1 + 2p_3)$-diagonal matrix $\text{diag}\{I_{p_1+p_3}, hI_{p_1}, bI_{p_3}\}$. Let $u_\vartheta(\vartheta, \beta) = E([\varphi\{\vartheta_0^T(t, U_i(t))\tilde{X}_i(t) + \beta_0^T X_{2i}(t)\} - \varphi\{\vartheta^T(t, U_i(t))\tilde{X}_i(t) + \beta^T X_{2i}(t)\}]\tilde{X}_i(t)\xi_i(t)\lambda_i(t)$ $|U_i(t) = u) f_U(t, u)$. Define $\vartheta_\beta^T(t, U_i(t))$ as the unique root such that $u_\vartheta(\vartheta, \beta) = 0$ for

$\beta \in \mathcal{N}_{\beta_0}$. Let

$$e_{\beta,11}(t,u) = E[\omega_i(t)\dot{\varphi}\{\vartheta_\beta^T(t,u)\tilde{X}_i(t) + \beta^T X_{2i}(t)\}\{\tilde{X}_i(t)\}^{\otimes 2}\xi_i(t)\lambda_i(t)]$$

and

$$e_{\beta,12}(t,u) = E[\omega_i(t)\dot{\varphi}\{\vartheta_\beta^T(t,u)\tilde{X}_i(t) + \beta^T X_{2i}(t)\}\tilde{X}_i(t)(X_{2i}(t))^T\xi_i(t)\lambda_i(t))].$$

When $\beta = \beta_0$, we have $\vartheta_\beta(t,u) = \vartheta_0(t,u)$. In this case, $e_{\beta,11}(t,u) = e_{11}(t,u)$ and $e_{\beta,12}(t,u) = e_{12}(t,u)$. □

Lemma B.3. *Under condition II, we have* $H\tilde{\vartheta}^*(t,u,\beta) \xrightarrow{\mathcal{P}} (\vartheta_\beta^{*T}(t,u), \mathbf{0}^T)^T$,

$$\partial H\tilde{\vartheta}(t,u,\beta)/\partial\beta \xrightarrow{\mathcal{P}} (-e_{\beta,12}(t,u)e_{\beta,11}^{-1}(t,u), \mathbf{0}^T)^T$$

uniformly in $t \in [t_1,t_2]$, $u \in [u_1,u_2]$ *and* β *in a neighborhood of* β_0.

Proof of Lemma B.3

To facilitate technical arguments, we will reparametrize the estimating function (3.2) via the transformation $\eta = H(\vartheta^* - \vartheta_0^*)$.

We first show that $\tilde{\eta} \to \mathbf{0}$ in probability, where $\tilde{\eta} = H(\hat{\vartheta}^* - \vartheta_0^*)$. Hence, the estimating function is

$$U_\vartheta(\eta,t_0,u_0) = \sum_{i=1}^n \int_0^\tau W_i(t)\left[Y_i(t) - \varphi\{\vartheta_\beta^{*T}\tilde{X}_i^*(t,t_0,u_0) + \eta^T H^{-1}\tilde{X}_i^*(t,t_0,u_0) + \beta^T X_{2i}(t)\}\right]$$
$$\times X_i^*(t,t_0,u_0)K_h(t-t_0)K_b(U_i(t)-u_0)\,dN_i(t).$$

By the Glivenko-Cantelli theorem and lemma B.1, and by exchanging the order of

expectation and integration, we have

$$
\begin{aligned}
& n^{-1}\{U(\eta,t_0,u_0)-U(0,t_0,u_0)\} \\
= \quad & -n^{-1}\sum_{i=1}^{n}\int_0^{\tau} W_i(t)\Big[\varphi\{\vartheta_\beta^{*T}\tilde{X}_i^*(t,t_0,u_0)+\eta^T H^{-1}\tilde{X}_i^*(t,t_0,u_0)+\beta^T X_{2i}(t)\} \\
& -\varphi\{\vartheta_\beta^{*T}\tilde{X}_i^*(t,t_0,u_0)+\beta^T X_{2i}(t)\}\Big]\tilde{X}_i^*(t,t_0,u_0)K_h(t-t_0)K_b(U_i(t)-u_0)\,dN_i(t) \\
\xrightarrow{\mathcal{P}} \quad & \int_{-1}^{1}\int_{-1}^{1} E\left[\omega_i(t_0)\dot{\mu}_i(t_0)\eta^T\begin{pmatrix}\tilde{X}_i(t_0)\\ X_{1i}(t_0)y\\ X_{3i}(t_0)x\end{pmatrix}\begin{pmatrix}\tilde{X}_i(t_0)\\ 0\end{pmatrix}\xi_i(t_0)\lambda_i(t_0)|U_i(t_0)=u_0\right] \\
& \times f_U(t_0,u_0)\,dy\,dx,
\end{aligned}
$$

uniformly in $t_0\in[t_1,t_2], u_0\in[u_1,u_2]$ and η in a neighborhood of 0.The limit has a unique root at $\eta=0$. By Lemma B.2 it follows that $\tilde{\eta}\to\mathbf{0}$ uniformly in t and u. Thus $H\tilde{\vartheta}^*(t,u,\beta)-(\vartheta_\beta^{*T}(t,u),0^T)^T\xrightarrow{\mathcal{P}}\mathbf{0}$ uniformly in t, u and $\beta\in\mathcal{N}_\beta$

Following the same steps as in Lemma A.1 in Chapter 2, and by using Lemma B.1 repeatedly, we can show that

$$\partial H\tilde{\vartheta}(t,u,\beta)/\partial\beta\xrightarrow{\mathcal{P}}(-e_{\beta,12}(t,u)e_{\beta,11}^{-1}(t,u),\mathbf{0}^T)^T$$

uniformly in $t\in[t_1,t_2]$, $u\in[u_1,u_2]$ and β in a neighborhood of β_0. □

Lemma B.4. *Under condition II, as $h\asymp b$, $nh^2\to\infty$ and $nh^6=O_p(1)$, we have*

$$
\begin{aligned}
& \sqrt{nhb}\{\tilde{\vartheta}(t_0,u_0,\beta_0)-\vartheta_0(t_0,u_0)-\frac{1}{2}h^2\nu_2 e_{11}^{-1}(t_0,u_0)b_\alpha(t_0,u_0)-\frac{1}{2}b^2\nu_2 e_{11}^{-1}(t_0,u_0)b_\gamma(t_0,u_0)\} \\
& =(e_{11}(t_0,u_0))^{-1}\sqrt{nhb}\mathbf{A}_n(t_0,u_0)+o_p(1) \qquad \text{(B.2)}
\end{aligned}
$$

uniformly in $t_0 \in [t_1, t_2]$ *and* $u_0 \in [u_1, u_2]$, *where*

$$\mathbf{A}_n(t_0, u_0) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau W_i(t)\{Y_i(t) - \mu_i(t)\}\tilde{X}_i(t)K_h(t - t_0)K_b(U_i(t) - u_0)\, dN_i(t).$$

Further, $\sqrt{nhb}\mathbf{A}_n(t_0, u_0) = O_p(1)$ *uniformly in* $t \in [t_1, t_2]$ *and* $u \in [u_1, u_2]$.

Proof of Lemma B.4

Because $U_\vartheta(\tilde{\vartheta}^*, t_0, u_0, \beta_0) = 0$, by Taylor expansion we have

$$H\{\tilde{\vartheta}^*(t_0, u_0, \beta_0) - \vartheta_0^*(t_0, u_0)\} = e_{11}^{-1}(t_0, u_0)\{n^{-1}H^{-1}U_\vartheta(\vartheta_0^*, t_0, u_0)\} + o_p(1) \quad \text{(B.3)}$$

The first p_1 components yields

$$\tilde{\vartheta}(t_0, u_0, \beta_0) - \vartheta_0(t_0, u_0) = e_{11}^{-1}(t_0, u_0)U_1(\vartheta_0^*, t_0, u_0) + o_p(1)$$

uniformly in t_0 and u_0, where

$$U_1(\vartheta_0^*, \beta_0) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau W_i(t)\left[Y_i(t) - \varphi\{\vartheta_0^{*T}\tilde{X}_i^*(t, t_0, u_0) + \beta_0^T X_{2i}(t)\}\right] \times \tilde{X}_{1i}(t)K_h(t - t_0)K_b(U_i(t) - u_0)\, dN_i(t).$$

Because by local linear approximation,

$$\begin{aligned}
&Y_i(t) - \varphi\{\vartheta_0^{*T}\tilde{X}_i^*(t, t_0, u_0) + \beta_0^T X_{2i}(t)\} \\
&= Y_i(t) - \mu_i(t) + \mu_i(t) - \varphi\{\vartheta_0^{*T}X_i^*(t, t_0, u_0) + \beta_0^T X_{2i}(t)\} \\
&= Y_i(t) - \mu_i(t) + \frac{1}{2}\dot{\mu}_i(t)\{(\ddot{\alpha}(t_0))^T X_{1i}(t)(t - t_0)^2 + (\ddot{\gamma}(u_0))^T X_{3i}(t)(U_i(t) - u_0)^2\} \\
&\quad + o_p((t - t_0)^2 + (U_i(t) - u_0)^2),
\end{aligned}$$

it yields $U_1(\vartheta_0^*, \beta_0) = \mathbf{A}_n(t_0, u_0) + \mathbf{B}_n(t_0, u_0) + \mathbf{C}_n(t_0, u_0) + o_p(h^2 + b^2)$, where

$$\mathbf{B}_n(t_0,u_0) = \frac{1}{2n}\sum_{i=1}^{n}\int_0^\tau W_i(t)\dot{\mu}_i(t)(\ddot{\alpha}(t_0))^T X_{1i}(t)(t-t_0)^2 \times \tilde{X}_{1i}(t)K_h(t-t_0)K_b(U_i(t)-u_0)\,dN_i(t), \tag{B.4}$$

and

$$\mathbf{C}_n(t_0,u_0) = \frac{1}{2n}\sum_{i=1}^{n}\int_0^\tau W_i(t)\dot{\mu}_i(t)(\ddot{\gamma}(u_0))^T X_{3i}(t)(U_i(t)-u_0)^2 \times\tilde{X}_{1i}(t)K_h(t-t_0)K_b(U_i(t)-u_0)\,dN_i(t) \tag{B.5}$$

By lemma 1, we conclude that

$$\begin{aligned}\frac{1}{h^2}\mathbf{B}_n(t_0,u_0) &= \frac{1}{2n}\sum_{i=1}^{n}\int_0^\tau W_i(t)\dot{\mu}_i(t)(\ddot{\alpha}(t_0))^T X_{1i}(t)\left(\frac{t-t_0}{h}\right)^2 \\ &\quad\times \tilde{X}_{1i}(t)K_h(t-t_0)K_b(U_i(t)-u_0)\,dN_i(t), \\ &\xrightarrow{\mathcal{P}} \frac{1}{2}\nu_2 b_\alpha(t_0,u_0),\end{aligned}$$

where $b_\alpha(t_0,u_0) = E\{\omega_i(t_0)\dot{\mu}_i(t_0)\tilde{X}_{1i}(t_0)X_{1i}^T(t_0)|U_i(t_0)=u_0\}f_U(t_0,u_0)\ddot{\alpha}(t_0)$, and

$$\begin{aligned}\frac{1}{b^2}\mathbf{C}_n(t_0,u_0) &= \frac{1}{2n}\sum_{i=1}^{n}\int_0^\tau W_i(t)\dot{\mu}_i(t)(\ddot{\gamma}(u_0))^T X_{3i}(t)\left(\frac{U_i(t)-u_0}{b}\right)^2 \\ &\quad\times \tilde{X}_{1i}(t)K_h(t-t_0)K_b(U_i(t)-u_0)\,dN_i(t) \\ &\xrightarrow{\mathcal{P}} \frac{1}{2}\nu_2 b_\gamma(t_0,u_0),\end{aligned}$$

where $b_\gamma(t_0,u_0) = E\{\omega_i(t_0)\dot{\mu}_i(t_0)\tilde{X}_{1i}(t_0)X_{3i}^T(t_0)|U_i(t_0)=u_0\}f_U(t_0,u_0)\ddot{\gamma}(u_0)$. Hence (B.2) holds. □

Proof of Theorems

Proof of Theorem 3.1

By Lemma B.1, Lemma B.3 and application of the Glivenko-Cantelli theorem to the estimating function defined in (3.4), we have

$$
\begin{aligned}
& n^{-1}U_\beta(\beta) \\
& \xrightarrow{\mathcal{P}} E\int_0^\tau W_i(t)[Y_i(t) - \varphi\{\vartheta_\beta^T(t, U_i(t))\tilde{X}_i(t) + \beta^T X_{2i}(t)\}] \\
& \qquad \times \{X_{2i}(t) - (e_{\beta,12}(t, U_i(t)))^T (e_{\beta,11}(t, U_i(t)))^{-1}\tilde{X}_i(t)\} dN_i(t) \\
& = E\int_0^\tau W_i(t)[\varphi\{\vartheta_0^T(t, U_i(t))\tilde{X}_i(t) + \beta_0^T X_{2i}(t)\} - \varphi\{\vartheta_\beta^T(t, U_i(t))\tilde{X}_i(t) + \beta^T X_{2i}(t)\}] \\
& \quad \times \{X_{2i}(t) - (e_{\beta,12}(t, U_i(t)))^T (e_{\beta,11}(t, U_i(t)))^{-1}\tilde{X}_i(t)\}\xi_i(t)\lambda_i(t)dt \\
& = u(\beta),
\end{aligned} \tag{B.6}
$$

where β_0 is the unique root of $u(\beta)$. Then by Theorem 5.9 of Van Der Vaart (1998), $\hat{\beta}\xrightarrow{\mathcal{P}}\beta_0$.

By Glivenko-Cantelli theorem and Lemma B.3,

$$
\begin{aligned}
& - n^{-1}\frac{\partial U_\beta(\beta)}{\partial \beta}|_{\beta=\beta_0} \\
& = n^{-1}\sum_{i=1}^n \int_0^\tau W_i(t)\dot{\varphi}\{\tilde{\vartheta}^T(t, U_i(t), \beta_0)\tilde{X}_i(t) + \beta_0^T X_{2i}(t)\} \\
& \qquad\qquad \times \left\{\frac{\partial\tilde{\vartheta}(t, U_i(t), \beta_0)}{\partial\beta}\tilde{X}_i(t) + X_{2i}(t)\right\} dN_i(t) + o_p(1) \\
& \xrightarrow{\mathcal{P}} E[\int_0^\tau \omega_i(t)\dot{\mu}_i(t)\{X_{2i}(t) - (e_{12}(t, U_i(t)))^T (e_{11}(t, U_i(t)))^{-1}\tilde{X}_i(t)\}^{\otimes 2} dN_i(t)] \equiv A_\beta
\end{aligned}
$$

Now we show that $n^{-1/2}U_\beta(\beta_0)$ converges in distribution to a normal distribution.

By Taylor expansion,

$$\begin{aligned}
&\varphi\{\tilde{\vartheta}^T(t,U_i(t),\beta_0)\tilde{X}_i(t)+\beta_0^T X_{2i}(t)\}-\varphi\{\vartheta_0^T(t,U_i(t))\tilde{X}_i(t)+\beta_0^T X_{2i}(t)\}\\
&=\dot{\mu}_i(t)\{\tilde{\vartheta}^T(t,U_i(t),\beta_0)-\vartheta_0^T(t,U_i(t))\}\tilde{X}_i(t)\\
&+O_p(\|\tilde{\vartheta}(t,U_i(t),\beta_0)-\vartheta_0(t,U_i(t))\|^2 \qquad \text{(B.7)}
\end{aligned}$$

By Lemma B.3 and Lemma B.4,

$$\begin{aligned}
&n^{-1/2}\sum_{i=1}^{n}\int_0^\tau W_i(t)[\varphi\{\tilde{\vartheta}^T(t,U_i(t),\beta_0)\tilde{X}_i(t)+\beta_0^T X_{2i}(t)\}\\
&\quad-\varphi\{\vartheta_0^T(t,U_i(t))\tilde{X}_i(t)+\beta_0^T X_{2i}(t)\}]\times\left\{(\tilde{X}_i(t))^T\frac{\partial\tilde{\vartheta}(t,U_i(t),\beta_0)}{\partial\beta}+(X_{2i}(t))^T\right\}dN_i(t)\\
&=n^{-1/2}\sum_{i=1}^{n}\int_0^\tau W_i(t)\dot{\mu}_i(t)\{\tilde{\vartheta}^T(t,U_i(t),\beta_0)-\vartheta_0^T(t,U_i(t))\}\\
&\quad\times\left\{(\tilde{X}_i(t)\tilde{X}_i(t))^T\frac{\partial\tilde{\vartheta}(t,U_i(t),\beta_0)}{\partial\beta}+\tilde{X}_i(t)(X_{2i}(t))^T\right\}dN_i(t)+o_p(1)\\
&=o_p(1). \qquad \text{(B.8)}
\end{aligned}$$

Hence,

$$\begin{aligned}
&n^{-1/2}U_\beta(\beta_0)\\
&=n^{-1/2}\sum_{i=1}^{n}\int_0^\tau W_i(t)\epsilon_i(t)\left\{\frac{\partial\tilde{\vartheta}(t,U_i(t),\beta_0)}{\partial\beta}\tilde{X}_i(t)+X_{2i}(t)\right\}dN_i(t)+o_p(1)\\
&=n^{-1/2}\sum_{i=1}^{n}\int_0^\tau W_i(t)\epsilon_i(t)\{X_{2i}(t)-(e_{12}(t,U_i(t)))^T(e_{11}(t,U_i(t)))^{-1}\tilde{X}_i(t)\}dN_i(t)+o_p(1)
\end{aligned}$$

which converges in distribution to a normal distribution with variance

$$\Sigma_\beta=E\left(\int_{t_1}^{t_2}\omega_i(t)\epsilon_i(t)\{X_{2i}(t)-(e_{12}(t,U_i(t)))^T(e_{11}(t,U_i(t)))^{-1}\tilde{X}_i(t)\}dN_i(t)\right)^{\otimes 2}.$$

Hence, $n^{1/2}(\hat{\beta}-\beta_0)\xrightarrow{\mathcal{D}}N(0,A_\beta^{-1}\Sigma_\beta A_\beta^{-1})$. □

Proof of Theorem 3.2

(a) Since $\hat{\vartheta}(t_0,u_0)=\tilde{\vartheta}(t_0,u_0,\hat{\beta})$, we have $\hat{\vartheta}(t_0,u_0)\xrightarrow{\mathcal{P}}\vartheta_0(t_0,u_0)$ uniform in $t\in[0,\tau]$ and $u\in[u_1,u_2]$ by Lemma B.1 and Theorem 1. Then

$$\begin{aligned}&\sup_{t_0\in[t_1,t_2]}|\hat{\vartheta}(t_0)-\vartheta_0(t_0)|=\sup_{t_0\in[t_1,t_2]}|n^{-1}\sum_{j=1}^{n}\{\hat{\vartheta}(t_0,U_j(t_0))-\vartheta_0(t_0,U_j(t_0))\}|\\ \leq\ &\sup_{t_0\in[t_1,t_2],u_0\in[u_1,u_2]}|\hat{\vartheta}(t_0,u_0)-\vartheta_0(t_0,u_0)|=o_p(1).\end{aligned}$$

(b) Following the proof of Lemma B.4, we obtain

$$\begin{aligned}&\sqrt{nhb}\{\tilde{\alpha}(t_0,u_0,\beta_0)-\alpha_0(t_0,u_0)\}\\ &=-\mathscr{I}_1 e_{11}^{-1}(t_0,u_0)\sqrt{\frac{hb}{n}}\sum_{i=1}^{n}\int_0^{\tau}W_i(t)\epsilon_i(t)\tilde{X}_{1i}(t)K_h(t-t_0)K_b(U_i(t)-u_0)\,dN_i(t)\\ &+\frac{1}{2}\sqrt{nhb}h^2\nu_2 e_{11}^{-1}(t_0,u_0)b_\alpha(t_0,u_0)+\frac{1}{2}\sqrt{nhb}b^2\nu_2 e_{11}^{-1}(t_0,u_0)b_\gamma(t_0,u_0)+o_p(\sqrt{nhb}(h^2+b^2))\end{aligned}$$

Note that $e_{11}^{-1}(t_0,u_0)b_\gamma(t_0,u_0)$ is zero for the first p_1 components and $e_{11}^{-1}(t_0,u_0)b_\alpha(t_0,u_0)$ is $\ddot{\alpha}(t_0)$ for the first p_1 components. Then

$$\begin{aligned}&\sqrt{nh}\{\hat{\alpha}(t_0)-\alpha_0(t_0)\}\\ &=-\sqrt{\frac{h}{n}}\sum_{i=1}^{n}\int_0^{\tau}W_i(t)\{Y_i(t)-\mu_i(t)\}\{n^{-1}\sum_{j=1}^{n}\mathscr{I}_1 e_{11}^{-1}(t_0,U_j(t_0))X_i(t)K_b(U_i(t)-U_j(t_0))\}\\ &\qquad\times K_h(t-t_0)\,dN_i(t)\\ &+\sqrt{\frac{h}{n}}n^{-1}\sum_{j=1}^{n}\{e_{11}(t_0,U_j(t_0))^{-1}e_{12}(t_0,U_j(t_0))\}(\hat{\beta}-\beta_0)\\ &+\frac{1}{2}\sqrt{nh}h^2\nu_2\ddot{\alpha}(t_0).\end{aligned}$$

By Lemma A.1 in Yin et al. (2008),

$$\frac{1}{n}\sum_{j=1}^{n} e_{11}^{-1}(t_0, U_j(t_0))K_b(u - U_j(t_0)) = e_{11}^{-1}(t_0, u) + O_p(\frac{\log b}{\sqrt{nb}}) + O(b^2)$$

uniformly in $t \in [t_1, t_2]$ and $u \in [u_1, u_2]$. It follows that

$$\begin{aligned}
&\sqrt{nh}\{\hat{\alpha}(t_0) - \alpha_0(t_0) - \frac{1}{2}h^2\nu_2\ddot{\alpha}(t_0)\} \\
&= \sqrt{\frac{h}{n}}\sum_{i=1}^{n}\int_0^\tau W_i(t)\left\{Y_i(t) - \mu_i(t)\right\}\mathscr{I}_1 e_{11}^{-1}(t_0, U_i(t))X_i(t)K_h(t - t_0)\, dN_i(t) + o_p(1) \\
&= n^{-1/2}\sum_{i=1}^{n} g_i(t_0) + o_p(1),
\end{aligned}$$

where $g_i(t_0) = h^{1/2}\int_0^\tau \left\{Y_i(t) - \mu_i(t)\right\}\mathscr{I}_1 e_{11}^{-1}(t_0, U_i(t))X_i(t)K_h(t-t_0)\, dN_i(t)$. Following the arguments of Lemma 2 of Sun (2010),

$$\sqrt{nh}(\hat{\alpha}(t) - \alpha_0(t) - \frac{1}{2}h^2\nu_2\ddot{\alpha}(t_0)) \xrightarrow{\mathcal{D}} N\left(0, \Sigma_\alpha(t)\right) \tag{B.9}$$

where

$$\Sigma_\alpha(t_0) = \lim_{n\to\infty} hE\left[\int_0^\tau \left\{Y_i(t) - \mu_i(t)\right\}\mathscr{I}_1 e_{11}^{-1}(t_0, U_i(t))X_i(t)K_h(t - t_0)\, dN_i(t)\right]^{\otimes 2}.$$

□

Proof of Theorem 3.3

Following the same argument as the proof of Theorem 3.2, we have $\hat{\gamma}(u) \xrightarrow{\mathcal{P}} \gamma_0(u)$ uniformly in $u \in [u_1, u_2]$, and $\sqrt{nb}(\hat{\gamma}(u) - \gamma_0(u) - \frac{1}{2}b^2\nu_2\ddot{\gamma}(u)) \xrightarrow{\mathcal{D}} N\left(0, \Sigma_\gamma(u)\right)$. □

Proof of Theorem 3.4

By Lemma B.4,

$$\begin{aligned}
&\sqrt{nh}\{\hat{\gamma}(u) - \gamma_0(u) - \frac{1}{2}b^2\nu_2\ddot{\gamma}(u)\} \\
&= \sqrt{nh}\{\tilde{\gamma}(u,\beta_0) - \gamma_0(u) - \frac{1}{2}b^2\nu_2\ddot{\gamma}(u)\} \\
&\quad - \sqrt{nh}E\{(e_{11}(U_j^{-1}(u),u))^{-1}e_{12}(U_j^{-1}(u),u)\}\{\hat{\beta}-\beta_0\}.
\end{aligned}$$

Thus

$$\begin{aligned}
G_n(u) = n^{-1/2}\sum_{i=1}^{n}\Big\{ &\int_{u_1}^{u}\int_0^{\tau}\omega_i(t)\epsilon_i(t)\mathscr{I}_3 e_{11}(t,s)^{-1}X_i(t)K_b(U_i(t)-s)\,dN_i(t)\,ds \\
&+ \int_{u_1}^{u}E\{(e_{11}(U_j^{-1}(s),s))^{-1}e_{12}(U_j^{-1}(s),s)\}\,ds A_\beta^{-1}\int_{t_1}^{t_2}\omega_i(t)[Y_i(t)-\mu_i(t)] \\
&\times\{X_{2i}(t) - (e_{12}(t,U_i(t)))^T(e_{11}(t,U_i(t)))^{-1}\tilde{X}_i(t)\}\,dN_i(t)\Big\} + o_p(1),
\end{aligned}$$

which converges weakly to a mean-zero Gaussian process by central limit theorem.

□